TURN-TAKING IN HUMAN COMMUNICATIVE INTERACTION

EDITED BY: Judith Holler, Kobin H. Kendrick, Marisa Casillas and Stephen C. Levinson

PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88919-825-2
DOI 10.3389/978-2-88919-825-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

TURN-TAKING IN HUMAN COMMUNICATIVE INTERACTION

Topic Editors:
Judith Holler, Max Planck Institute for Psycholinguistics, Netherlands
Kobin H. Kendrick, Max Planck Institute for Psycholinguistics, Netherlands
Marisa Casillas, Max Planck Institute for Psycholinguistics, Netherlands
Stephen C. Levinson, Max Planck Institute for Psycholinguistics, Netherlands

Cover image: Amsterdam, The Netherlands, traffic lights for cyclists.
Image used under license from Shutterstock.com

The core use of language is in face-to-face conversation. This is characterized by rapid turn-taking. This turn-taking poses a number central puzzles for the psychology of language. Consider, for example, that in large corpora the gap between turns is on the order of 100 to 300 ms, but the latencies involved in language production require minimally between 600 ms (for a single word) or 1500 ms (for as simple sentence). This implies that participants in conversation are predicting the ends of the incoming turn and preparing in advance. But how is this done? What aspects of this prediction are done when? What happens when the prediction is wrong? What stops participants coming in too early? If the system is running on prediction, why is there consistently a mode of 100 to 300 ms in response time?

The timing puzzle raises further puzzles: it seems that comprehension must run parallel with the preparation for production, but it has been presumed that there are strict cognitive limitations on more than one central process running at a time. How is this bottleneck overcome? Far from being 'easy' as some psychologists have suggested, conversation may be one of the most demanding cognitive tasks in our everyday lives. Further questions naturally arise: how do children learn to master this demanding task, and what is the developmental trajectory in this domain?

Research shows that aspects of turn-taking, such as its timing, are remarkably stable across languages and cultures, but the word order of languages varies enormously. How then does

prediction of the incoming turn work when the verb (often the informational nugget in a clause) is at the end? Conversely, how can production work fast enough in languages that have the verb at the beginning, thereby requiring early planning of the whole clause? What happens when one changes modality, as in sign languages – with the loss of channel constraints is turn-taking much freer? And what about face-to-face communication amongst hearing individuals – do gestures, gaze, and other body behaviors facilitate turn-taking? One can also ask the phylogenetic question: how did such a system evolve? There seem to be parallels (analogies) in duetting bird species, and in a variety of monkey species, but there is little evidence of anything like this among the great apes.

All this constitutes a neglected set of problems at the heart of the psychology of language and of the language sciences. This Research Topic contributes to advancing our understanding of these problems by summarizing recent work from psycholinguists, developmental psychologists, students of dialog and conversation analysis, linguists, phoneticians, and comparative ethologists.

Citation: Holler, J., Kendrick, K, H., Casillas, M., Levinson, S. C., eds. (2016). Turn-Taking in Human Communicative Interaction. Lausanne: Frontiers Media.
doi: 10.3389/978-2-88919-825-2

Table of Contents

5. Turn-taking in signed languages

6. Development of turn-taking skills

frontiers in Psychology

EDITORIAL
published: 21 December 2015
doi: 10.3389/fpsyg.2015.01919

Editorial: Turn-Taking in Human Communicative Interaction

Judith Holler, Kobin H. Kendrick, Marisa Casillas and Stephen C. Levinson*

Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

Keywords: turn-taking, turn-timing, turn transitions, conversation, social interaction, language processing, prediction, projection

The Editorial on the research topic

Turn-Taking in Human Communicative Interaction

One intriguing feature of the human communication system is the interactional infrastructure it builds on. In both dyadic and multi-person interactions, conversation is highly structured and organized according to set principles (Sacks et al., 1974). Human adult interaction is characterized by a mechanism of exchange based on alternating (and relatively short) bursts of information. In the majority of cases, only one person tends to speak at a time and each contribution usually receives a response. What is remarkable is the precise timing of these sequential contributions, resulting in gaps between speaking turns averaging around just 200 ms (Stivers et al., 2009). From psycholinguistic experiments, we know that the time it takes to produce even simple one-word-utterances (min. 600 ms, Indefrey and Levelt, 2004) by far exceeds this average gap duration, hinting at the complexity of the cognitive processes that must be involved (Levinson, 2013).

While the behavioral principles governing turn-taking in interaction have been researched for some decades—primarily by scholars of conversation analysis—the cognitive underpinnings of the human turn-taking system have long remained elusive. Recently, psycholinguists have begun to explore the cognitive and neural processes that allow us to deal effectively with the immensely complex task of taking turns on time. Amongst other things, this has highlighted the anticipatory, predictive processes that must be at work, as well as the different layers of processing allowing production planning and comprehension to take place simultaneously (de Ruiter et al., 2006; Magyari and de Ruiter; Bögels et al., 2015). These insights mesh well with the conversation analytic literature that has illuminated the interactional environments in which individual turns are embedded: their sequential organization and the use of conventionalized linguistic constructions allow for the projection of upcoming talk, as well as for the recognition of points of possible completions in the turn which make transition to the next speaker relevant (Sacks et al., 1974; Ford and Thompson, 1996; Schegloff, 2007). The articles in this Research Topic bring together these as yet largely independent lines of research to elucidate our understanding of turn-taking from multiple perspectives and aim to foster future synergies.

In addition to exploring the adult psycholinguistic machinery and its workings, researchers have begun to wonder how and when the required cognitive and social processes mature in children, as well as how they compare to those in other species. Levinson (2006) proposed that human beings are inherently social and interactive in orientation. He argues that an "interaction engine" may lie at the heart of children's early predisposition for turn-taking. Likewise, this particular human capacity might explain the strong cultural universals in the structure of human interaction as well as the striking commonalities and differences in communication systems brought about by the course of evolution.

The present Research Topic provides a collection of experimental and observational empirical studies using qualitative and quantitative approaches, complemented by articles offering reviews,

OPEN ACCESS

Edited and reviewed by:
Manuel Carreiras, Basque Center on Cognition, Brain and Language, Spain

****Correspondence:***
Judith Holler
judith.holler@mpi.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *26 November 2015*
Accepted: *30 November 2015*
Published: *21 December 2015*

Citation:
Holler J, Kendrick KH, Casillas M and Levinson SC (2015) Editorial: Turn-Taking in Human Communicative Interaction. Front. Psychol. 6:1919. doi: 10.3389/fpsyg.2015.01919

opinions, and models. They aim to inform the reader about the most recent advances in our endeavor of unraveling the workings of the human turn-taking system in communicative interaction. The contributions are organized into six sections: (1) Foundations of turn-taking, (2) Signals and mechanisms for prediction and timing, (3) Planning next turns in conversation, (4) Effects of context and function on timing, (5) Turn-taking in signed languages, and (6) Development of turn-taking skills.

FOUNDATIONS OF TURN-TAKING

The articles in this section outline models of human turn-taking, specify the interaction of the various psycholinguistic processes that underlie our ability to take conversational turns on time, and test the applicability of human turn-taking models to non-human animal species. Levinson and Torreira review behavioral and cognitive findings specifying the parameters of the processes underlying the human turn-taking system. This empirical evidence is synthesized into a model claiming that intention ascription and response planning begin as early as possible during the incoming turn, running through all the serial stages of speech production à la Levelt (1989) before the response is launched, triggered by turn-final cues. Garrod and Pickering propose a model that specifies two processes. The first is based on the entrainment of brain oscillations that allow listeners to predict when the incoming turn will end. The second is constrained by the first and based on covert imitation, allowing listeners to determine the intention conveyed by the incoming turn. The final article in this section addresses the phylogenetic development of turn-taking skills. Henry et al. look at the European Starling's turn-taking behavior, finding evidence for both temporal and structural regularities, the influence of the immediate as well as the wider social context in which turns are produced, and of emitter-specific factors influencing the behavior—thus pointing toward strong similarities with some of the features shaping turn-taking in humans. In addition, they provide comparisons with other starling species, leading the authors to argue for turn-taking behavior having co-evolved in close interdependency with social structure.

The empirical studies collected in the rest of this Research Topic support various components of these proposed turn-taking models while in places being at odds with some of the claims made. As much as the current volume is a summary of the state-of-the-art in the field, it also aims to stimulate future research that will help us piece together the parts of the remarkable puzzle that human turn-taking poses.

SIGNALS AND MECHANISMS FOR PREDICTION AND TIMING

One of the central debates on the cognitive processes involved in turn-taking focuses on the role played by prediction. Part of this debate is the issue of which kinds of cues adults may use for predicting the end of turns, allowing them to come in on time. The article by Riest et al. further advances this debate by testing, in three offline experiments, the relative contribution of syntactic, and semantic information to turn-end anticipation. It shows that, while both types of information are essential, adults rely predominantly on the latter. The article by Holler and Kendrick builds on this work by using eye-tracking technology to investigate the responses of observers directly immersed in a conversational setting. The data show that observers' eye movements toward next speakers are not random but guided by points of possible completion in current turns, thus revealing interactants' sensitivity and orientation toward the semantic, syntactic, prosodic, and pragmatic information that becomes available as turns unfold. The article by Hiroko zooms into the projective power of specific lexicogrammatical particles in Japanese (*wa*, *mo,* and *tte*). These become available to listeners as turns unfold in conversation and often allow next speakers to predict the content of ongoing turns. Himbert et al. throw light on yet another source of information that speakers in interaction may use for timing their turns: their analysis demonstrates that interlocutors adapt their turn-taking rhythms to one another, which they argue is facilitated by the alignment of semantic and syntactic processes.

PLANNING NEXT TURNS IN CONVERSATION

The contributions in this section explore some of the cognitive processes involved in preparing next turns in conversation. Applying a cutting edge statistical approach ("random forests") to data from a large conversational corpus, Roberts et al. explore the value of both psycholinguistic factors (e.g., word frequency and syntactic complexity) and conversational structures (e.g., the sequential relationships between turns) as explanatory factors when modeling the timing of turns in conversation. Their results show that both sets of factors significantly contribute to explaining variation in turn timing. Torreira et al. study pre-answer in-breaths in a dialogue setting using insights from acoustic and inductive plethysmography recordings. They demonstrate that the occurrence of an in-breath is dependent on the length of an answer, suggesting that answers are planned prior to these in-breaths. Since the pre-answer in-breaths in their data were launched close to the end of question turns, the data provide evidence for the concurrence of comprehension and next utterance planning.

EFFECTS OF CONTEXT AND FUNCTION ON TIMING

Three articles investigate the interplay of turn-taking rules with other principles shaping human behavior in specific conversational contexts. Kendrick shows that turns dealing with problems of speaking, hearing, and understanding (i.e., other-initiations of repair) are governed by different timing principles and can thus break the common pattern of minimal gaps between turns. As the analysis reveals, the longer gaps characteristic of repair sequences tend to be used by participants as opportunities to either allow the producer of the trouble source to resolve the issue before repair is initiated, to allow

themselves to resolve their problems in understanding before initiating repair, or to signal problems in understanding through visual displays (e.g., eyebrow raise) before initiating repair verbally. The article by Gardner and Mushin provides evidence from Garrwa, an indigenous Australian language, for turns that are followed by substantially longer gaps than one would ordinarily expect based on prior work on English conversations. In these cases, however, it is not repair that drives the longer turn transition times; the environment in which they occur is slow-paced conversation, appearing to reduce the pressure for gap minimization. Stevanovic and Peräkylä discuss perspectives on the intersection of two different systems of temporal organization, that of turns at talk and that of emotional reciprocity—the former favoring sequential organization, the latter affording simultaneity and immediate uptake through emotional contagion and mimicry.

TURN-TAKING IN SIGNED LANGUAGES

The research presented in this section investigates the principles of turn-taking and sequence organization in signed languages where communication is constrained to the visual modality. De Vos et al. analyze the timing of turns in Sign Language of the Netherlands (NGT), showing that the timing of turns in signed conversation looks remarkably similar to that of spoken interaction (i.e., with minimal gaps and minimal overlaps) when considering not simply onset and offset of manual movements but individual movement phases (preparations, strokes, retractions). Girard-Groeber examines turn-taking principles in multi-party conversations in Swiss German Sign Language (DSGS), focusing on the occurrence of overlaps. She, too, finds striking similarities with spoken interactions: the examples provided illustrate a strong orientation to the "one at a time" principle, an orientation of participants toward points of possible completion in the sign stream, and a set of principles that appear to determine deviations from this rule (such as repair initiations or strong disagreements). Manrique and Enfield focus on a particular type of turn transition environment—other-initiated-repair—in Argentine Sign Language (LSA), thus complementing Kendrick's work on repair in spoken interaction (this volume). However, their focus is on how repair is elicited in visual question-answer sequences rather than on the timing of turns in the repair environment, revealing the frequent use of a visual display form termed the "freeze-look." Next to clearly unique features, the three articles point toward some striking similarities regarding the timing and organization of turns in spoken and signed languages.

DEVELOPMENT OF TURN-TAKING SKILLS

Convergent findings regarding principles governing turn-taking across languages in different modalities hint at the possibility of a shared cognitive infrastructure underlying all human communicative interaction. This cognitive infrastructure may also account for the ease with which young children appear to acquire the necessary skills to interact with others. The contributions included in this section focus on the acquisition of turn-taking in very young infants and in children as they start to master spoken language. The first two articles suggest that temporal turn-taking skills are learned early on in infancy. Gratier et al. demonstrate that already at 8–21 weeks babies are active participants in, as well as initiators of, turn-taking sequences, but also that at this early stage of development mothers play a core role in the timing of turns by adapting their behavior to the infant. Hilbrink et al. provide a longitudinal study showing that turn-timing skills continue to develop continuously from 3 to 18 months, with some regressive slowing down as language comprehension kicks in around the "9 month revolution" (Tomasello, 2008). Clark and Lindsey provide a case study of one child's (1;4-3;5 years) verbal and gestural responses to questions. The pattern they find nicely fits with the temporal slowing down in vocal turn-timing caused by the challenge of having to master language—while verbal responses often occurred with long delays, the child frequently produced gestural responses preceding speech. The following two articles examine children's use of linguistic cues for anticipating upcoming next turns when observing dyadic conversations. Keitel and Daum find that three but not 1 year olds are able to make use of intonational cues for predicting upcoming next turns. In line with this, Lammertink et al. find that 2 year olds make use of prosodic cues for predicting upcoming next turns, but that they make use of lexicosyntactic cues, too, even weighing these more strongly—just like adults do.

FUNDING

The authors were supported through the Max Planck Gesellschaft and European Research Council (Advanced grant #269484 INTERACT awarded to SCL) during the preparation of the editorial and the research topic as a whole.

REFERENCES

Bögels, S., Magyari, L., and Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Sci. Rep.* 5:12881. doi: 10.1038/srep12881

de Ruiter, J., Mitterer, H., and Enfield, N. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

Ford, C. E., and Thompson, S. A. (1996). "Interactional units in conversation: syntactic, intonational, and pragmatic resources for the projection of turn completion," in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 135–184.

Indefrey, P., and Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levinson, S. C. (2006). "On the human "interaction engine"," in *Roots of Human Sociality: Culture, Cognition and Interaction,* eds N. J. Enfield and S. C. Levinson (Oxford: Berg), 39–69.

Levinson, S. C. (2013). "Action formation and ascription," in *The Handbook of Conversation Analysis,* eds T. Stivers and J. Sidnell (Malden, MA: Wiley-Blackwell), 103–130.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Schegloff, E. A. (2007). *Sequence Organization in Interaction: Volume 1: A Primer in Conversation Analysis.* Cambridge: Cambridge University Press.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 June 2015
doi: 10.3389/fpsyg.2015.00731

Timing in turn-taking and its implications for processing models of language

Stephen C. Levinson[1,2]* *and Francisco Torreira*[1]

[1] Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

The core niche for language use is in verbal interaction, involving the rapid exchange of turns at talking. This paper reviews the extensive literature about this system, adding new statistical analyses of behavioral data where they have been missing, demonstrating that turn-taking has the systematic properties originally noted by Sacks et al. (1974; hereafter SSJ). This system poses some significant puzzles for current theories of language processing: the gaps between turns are short (of the order of 200 ms), but the latencies involved in language production are much longer (over 600 ms). This seems to imply that participants in conversation must predict (or 'project' as SSJ have it) the end of the current speaker's turn in order to prepare their response in advance. This in turn implies some overlap between production and comprehension despite their use of common processing resources. Collecting together what is known behaviorally and experimentally about the system, the space for systematic explanations of language processing for conversation can be significantly narrowed, and we sketch some first model of the mental processes involved for the participant preparing to speak next.

Keywords: turn-taking, conversation, language processing, language production, language comprehension

OPEN ACCESS

Edited by:
Manuel Carreiras, Basque Center on Cognition, Brain and Language, Spain

Reviewed by:
Brian MacWhinney, Carnegie Mellon University, USA
Martin John Pickering, The University of Edinburgh, UK

***Correspondence:**
Stephen C. Levinson, Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, Netherlands
stephen.levinson@mpi.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *28 January 2015*
Accepted: *16 May 2015*
Published: *12 June 2015*

Citation:
Levinson SC and Torreira F (2015) Timing in turn-taking and its implications for processing models of language. Front. Psychol. 6:731. doi: 10.3389/fpsyg.2015.00731

1. Introduction: Why Turn-Taking in Conversation is Important for the Psychology of Language

One of the most distinctive ethological properties of humans is that they spend considerable hours in the day in a close (often face-to-face) position with others, exchanging short bursts of sound in a human-specific communication pattern: extrapolating from Mehl et al. (2007), we may each produce about 1200 of these bursts a day, for a total of 2–3 h of speech. The bursts tend to involve a phrasal or clausal unit, but can be longer or shorter. At the end of such bursts, a speaker stops, and another takes a turn. This is the prime ecological niche for language, the context in which language is learned (see Section 6.1 below), in which the cultural forms of language have evolved, and where the bulk of language usage happens.

This core form of language use poses a central puzzle for psycholinguistics (see Section 6), which has largely ignored this context, instead examining details of the processes of language production or comprehension separately in laboratory contexts. Yet this prime use of language involves rapid switching between comprehension and production at a rate implying that these processes must sometimes overlap. Decades of experimentation have shown that the language production system has latencies of around 600 ms and up for encoding a new word (reviewed in Section 6.3) but the gaps between turns average

around 200 ms (see Section 5). This would seem to imply that participants planning to respond are already encoding their responses while the incoming turn from the other speaker is still unfinished. This in turn implies potentially long-range prediction in comprehension. A sketch model of the interleaving of comprehension and production processes is presented in Section 7.

To appreciate the full nature of this puzzle, it is essential to review what we know about the turn-taking system and its temporal properties. In Section 2, we review the foundational Sacks et al. (1974; henceforth SSJ) model of turn-taking, considering alternative proposals in Sections 3 and 4. The model proposes extensive prediction (or 'projection') of turn-ends, and an expectation of swift response. The systematicity of turn-taking and its temporal patterning are borne out by extensive corpus analyses (Section 5). We then turn to the psycholinguistic literature (Section 6), noting that sensitivity to turn-end cues is already shown early in child development. We point out that there is considerable evidence for predictive language comprehension, and for long latencies in language production, so that the central psycholinguistic puzzle (Section 6.5) posed by turn-taking seems to be resolved by predicting what the other interlocutor is going to say. Some direct recent investigations seem to bear this out (Section 6.4), although experimentation in this field is in its infancy. In Section 7 we take stock of the recent findings, and sketch a processing model addressing some of the processing puzzles involved.

2. Turn-Taking as a System: Research from Conversation Analysis

Sacks et al. (1974; SSJ) initiated the modern literature on conversational turn-taking by outlining how this behavior constitutes a system of social interaction with specific properties. It is not organized in advance (by say an order of speaking, or set units to be uttered), but is highly flexible, allowing for longer units when so mutually arranged, and organizing an indeterminate number of participants into a single conversation. The authors note that "overwhelmingly one speaker talks at a time. Occurrences of more than one speaker at a time are common but brief [. . .] Transitions (from one turn to the next) with no gap and no overlap are common, and together with slight gaps and slight overlaps make up the majority of transitions" (Sacks et al., 1974, p. 700). Obviously, such turn-taking behavior contrasts with the absence of turn-taking in cheering, heckling, laughing, etc. That things could be otherwise in the speech domain is shown by the contrasting speech exchange systems we also use, as in lectures where questions come at the end, or in a press conference where questions come from many parties but are answered by one, contrasting with a classroom where questions may come from the teacher alone, and may be answered by many volunteers. The importance of the conversational system is that, unlike the others, it appears to be the default mode of language use, as shown by its operation in the context of language learning, and among friends and family. As far as we know, it operates in a strongly universal way (cf. Stivers et al., 2009, 2010), while the other speech exchange systems are mostly culture-specific.

Sacks et al. (1974) argued that conversation is an elemental piece of social organization, regulated by social norms that prescribe one speaker at a time but allow open participation. The model they suggested consists of turn units and rules that operate over those units. The units they suggested are variable sizes of syntactic units, whose functions as full turns can be indicated prosodically. The end of such a unit constitutes a 'transition relevance place' or TRP. The rules specify:

(1) If the current speaker C selects the next speaker N, then C must stop, and N should start. ('Selection' could involve address terms, gaze, or in the case of dyadic conversation defaults to the other.)
(2) If C does not select N, than any participant can self-select, first starter gaining rights to that next unit.
(3) If no other party self-selects, C may continue.

These rules then recursively apply at each TRP.

These rules predict that intra-speaker silent gaps (generated by rule 3) will be longer than inter-speaker ones, a fact shown to be correct on large samples of conversation [ten Bosch et al., 2005 report gaps between continuations by the same speaker to be about 140 ms (c. 25%) longer than the average gap in turn transitions between different speakers]. It has also been suggested that on this basis a turn-taking 'beat' or 'clock' (with a period between 80 and 180 ms) can be discerned, suggesting a model of coupled oscillators that allow participants to synch (Wilson and Zimmerman, 1986; Wilson and Wilson, 2005).

It was evident to Sacks et al. (1974) that the model had consequences for language processing. They noted that, given that interlocutors may be addressed at any point, the system enforces obligate listening. More importantly, they noted that the speed of speaker transition would require 'projection' (prediction) of the end of the incoming turn, and production processes would have to begin before the end of the incoming turn, in part because turn beginnings have to be designed to facilitate that very projection (Sacks et al., 1974, 719; Levinson, 2013). Later corpus studies have established, as we shall see (see Section 4), that the great proportion of turn transitions fall between −100 and 500 ms, that is, between a short stretch of overlap to a gap with a duration equivalent to one to three syllables.

There is a great deal of later work in conversation analysis (CA) that has contributed to our understanding of this system (see Clayman, 2013; Drew, 2013; Hayashi, 2013 for overviews). It is important to appreciate that not all overlapping of turns can be understood as behavior that violates the rules above – some authors (see Section 4) have seen the frequency of overlap as undercutting the Sacks et al. (1974) model. Sacks et al. (1974) claimed that overlaps are common, but usually very short, and often accounted for by little additions to the first turn like address

forms or tags [as in (1)], or by misanalyses of when the turn is coming to an end [as in (2) where 'biscuits' was projected as the turn-end but it was followed by 'and cheese'; overlap indicated with square brackets]:

(1) Sacks et al. (1974, p. 707)

(9) A: Uh you been down here before [havenche.
B: Yeh. [NB: III:3:5]

(2) Jefferson (1984, p. 15)

1. Vera: they muucked intuh biscuits. They had (.) quite a lotta
2. –> biscuit [s'n ch] e e | : : : s e.]
3. Jenny: –> [Oh : :] well thaht's it th]en [ye[s.

Note especially that some overlaps – namely competing (more or less simultaneous) first starts – are expectable by the rules above (as when two people start simultaneously by rule 2, or a participant operating rule 2 is a bit slow and overlaps with the current speaker continuing by rule 3). In these cases one or the other of the speakers normally drops out (impressionistic gap duration in seconds between brackets):

(3) Hayashi, 2013, p. 176 (from Auto Discussion)

(1) Curt: Mmm I'd like t'get a, high one if I cou:ld.
(2) (0.7)
(3) Gary: –> [I know uh-]
(4) Mike: –> [Lemme ask] a guy at work. He's gotta bunch a' old clu[nkers.

When there is competition to maintain the floor in these and other cases, this is often negotiated on a syllable by syllable basis, with e.g., deceleration, increase of intensity, and repeated syllables or words, until one speaker drops out (Schegloff, 2000).

Just as different kinds of overlap can be discerned, so can different kinds of absence of speech, differentiating between pauses (e.g., between units by the same speaker), gaps (between speakers), silences (meaningful absence of speech, e.g., after a question), and lapses (where no-one has self-selected to speak). It has been suggested (citations below) that participants are very sensitive to timing, so that an excessively long gap after a question, for instance, may be taken to indicate that the recipient has some kind of problem with it, for example finding it difficult to answer in the affirmative, or has uncertainty about the response. In the following a telephone caller takes gap of around 2 s to indicate the answer 'no,' which he himself then pre-emptively provides:

(4) Levinson, 1983, p. 320

C: So I was wondering would you be in your office on Monday (.) by any chance?
(2.0 s)
C: Probably not.

A considerable body of work has gone into understanding the role of extended gaps or silences in 'dispreferred' responses (responses not in line with the suggested action in the prior turn; see Pomerantz and Heritage, 2013 for review). Corpus analysis shows that gaps of 700 ms or more are associated with dispreferred actions, and that gaps longer than the norm (>300 ms) decrease the likelihood of an unqualified acceptance, and increase the likelihood that a response, be it acceptance or rejection, will have a dispreferred turn format (e.g., *Yes, but*. . . in the cases of acceptances; Kendrick and Torreira, 2015). Experimental work also shows that gaps of 600 ms or longer generate inferences of this unwelcome kind (Roberts et al., 2011).

The CA approach to turn-taking raises two major issues. The first is what exactly counts as a turn, and how participants can recognize such a unit as complete. The problem is that just about any word or phrase may in context constitute a turn, while syntactic units can be nested or conjoined indefinitely. Regarding this issue, Sacks et al. (1974, p. 721) note that "some understanding of sound production (i.e., phonology, intonation, etc.) is also very important to turn-taking organization." Thus in the following (drawn from the discussion in Clayman, 2013, p. 155), the terminal intonation contours do not occur till the end of the turns, and two turns each composed of three possibly complete syntactic units (divided by §) occur uninterrupted (note the whole is recognized by the recipient as a story under way, hence the continuers, which are themselves possibly elicited by rising intonation marked with '?'):

(5) Ford and Thompson (1996, p. 151)
K: Vera (.) was talking §on the phone §to her mom?
(6) C: mm hm
K: And uh she got off §the phone §and she was incredibly upset?
C: Mm hm.

In addition to syntactic and prosodic completeness, pragmatic completeness may be required to terminate a turn (Ford and Thompson, 1996; Levinson, 2013). Clearly a responsive action following the first part of a pair of actions like questions and answers, offers and acceptances, requests and compliances can be inspected for pragmatic efficacy; elsewhere the larger role in a sequence of speech acts may need to be satisfied.

The second major issue is 'projection' or predictive language understanding. Sacks et al. (1974) thought it clear that the turn-taking system can only work if there is extensive prediction in comprehension, so that recipients can use the unfolding turn to project an overall syntactic and prosodic envelope which would allow them to foresee when and how a turn would come to an end (see Clayman, 2013 for a review). It is not at all clear how this works, given the flexibility and extendibility of most syntactic units. Still, interesting insights are provided by such phenomena as turn-completion by the other, studied in depth by Lerner (1991, 2002; see also Hayashi, 2013). A typical example is where a bi-clausal structure is begun by speaker A, and the second clause completed by speaker B as below. Clearly an *If..then..* or *Whenever. . ., X. . .* structure projects a second downstream clause.

(7) Lerner (1991, p. 445)

1. Rich: if you bring it intuh them
2. Carol: –> ih don't cost yuh nothing.

Such cases do not alone show that recipients accurately predict the content of the second clause (indeed sometimes a jokey exploitation of the structure may appear). But sometimes exactly the same words do occur in overlap:

(8) Lerner (1991, p. 239)

1A: You have too many white friends. You don't know
2 how to be with (.) your p[eople
3B: -> [people
4A: Why are you not proud of (0.2) [you:r [peo:ple
5C: -> [you:r [peo:ple
6B: -> [°people°.

Such intrusions into others' turns are rare, and can act as demonstrations of understanding, occurring more routinely if speaker A is obviously engaged in a word-search and speaker B can provide the item. Cases like these demonstrate that extensive projection is possible, and psycholinguistic evidence supports this (Predictive Language Comprehension).

3. An Alternative Signaling Approach

The term 'turn-taking' was independently suggested by Yngve (1970) and Duncan (1972). Contemporaneous with the approach by Sacks et al. (1974), Duncan (1972, 1974) proposed, using videotapes of dialogs, a set of turn-taking signals. The main set are turn-handing-over signals, and consists of half a dozen cues: prosodic (type of final intonation, final syllable duration, final drop in pitch, or loudness), gestural (end of a gesture), and lexical/syntactic (tag, clause end, etc.). A second proposed signal is turn-maintaining and consists of a final mid-tone, continuing gesture or a gaze switch away. Turns followed by speaker change were found to nearly always occur with one or more turn-ending cues. On this basis, Duncan advanced a model where the turn-taking system is entirely under the control of the current speaker. This contrasts with the CA model, where speaker transition is contingently achieved by one speaker coming to the end of a unit and another starting (e.g., by self-selection). In addition, in the CA model there are no context-free signals: e.g., in English, a final mid tone usually marks turn-holding, but in specific contexts it may indicate turn-yielding (as when the conjunction *or* is appended to polar questions, e.g., *Are you leaving, or. . .?*); thus turn-taking can only be achieved on some much more global understanding of the incoming turn.

Although the signaling view is largely superseded, the research drew attention to (a) the importance of visual cues, and (b) the coincidence of turn transitions with a number of features of turn construction, prosody, gesture, etc. Kendon (1967) had earlier described different patterns of gaze between speakers (who alternately look away and look to addressees) and addressees (who gaze longer at the speaker). Goodwin (1980) later proposed a rule that sometime during the course of a turn a speaker should glance at the addressee, expecting to find a gazing addressee whenever he or she looks. The idea that speaker gaze when returning to addressee could function as a turn-yielding cue is, however, not easy to substantiate; More recently, Rossano (2013) has suggested this is because gaze is actually oriented to larger units of conversation (sequences), which it may serve to open and close.

4. Challenges to the Standard Model

Recently an alternative view to the Sacks et al. (1974) account was advanced by Heldner and Edlund (2010), who argue that turn-taking does not have all of the systematic properties described by Sacks et al. (1974). First, they find fault with the claim that speakers aim at no gap and no overlap. Actual zero gaps (under 10 ms) represent less than 1% of transitions and overlaps average 40% of transitions in their corpora. "From these observations, we conclude that the target with respect to timing of turn-taking cannot be one-speaker-at-a-time and no-gap–no-overlap, and furthermore that precision timing in turn-taking can neither be used in arguments in favor of projection, nor against reaction as models of timing in turn-taking" (Heldner and Edlund, 2010, p. 567). We believe these conclusions are misguided, and spell out the reasons here.

First, a target of 10 ms precision may not be realistic of human performance. Voiceless stops in English average between 60 and 80 ms (Crystal and House, 1988; Byrd, 1993), and at the end of a turn will be hard to distinguish from the beginning of the gap. Perceptual "no gap" was always estimated by conversation analysts to be of the order of 150–250 ms (i.e., close to the speaker transition mode; Schegloff, 2000). Heldner (2011) himself has gone on to show most usefully that a gap or overlap under 120 ms is not perceived as gap or overlap, respectively. It is interesting to compare the tolerable degree of lag in cross-modal matching as in the McGurk effect: an auditory signal following a visual one by up to 180 ms will still seem to be synchronized (Munhall et al., 1996). The majority (51–55%) of all turn transitions across corpora take place in under 200 ms (Heldner and Edlund, 2010, p. 563).

Second, as explained above, overlaps are of different kinds, some (e.g., continuers like *hmhm*, or minimal terminal overlaps) not being heard as intrusions on the turn, and others (like competing first starts) being specifically expectable. Below we provide a quantitative study of overlap (Overlap), which shows that overlaps tend to be minimal in size and occupy less than 5% of the speech stream.

Meanwhile, the argument that there is no target to avoid overlap seems unlikely. Qualitative analysis shows, as mentioned, that when overlap occurs, one speaker tends to rapidly drop out [as in example (3) above] so that the bulk of overlaps are of short duration. 'Interruption' is a sanctionable breach of social mores, as every child learns. The systematic properties of all the corpora that have been studied would be entirely different if overlap was not avoided.

On the basis of their dismissal of the no-gap–no-overlap target, Heldner and Edlund (2010, p. 566) go on to attack

further aspects of the standard model: "Thus, the no-gap–no-overlap principle (Sacks et al., 1974) can neither be used as a part of an argument in favor of projection nor against reaction simply because the no-gap–no-overlap cases hardly ever occur in real speaker change data. Importantly, this means that a principal motivation for projection in turn-taking is invalid." This attack on projection as a central element of the model will prove misplaced when we turn to consider the psycholinguistic evidence below (in fact Heldner and Edlund, 2010, p. 566 later concede that projection of content may be responsible for overlaps and short gaps).

The central plank of the dismissal of projection is that turn-taking is often not as rapid as has been claimed. Heldner and Edlund (2010, p. 563) note:

> "The cumulative distribution above the 200 ms threshold was also of interest, as it represented the cases where reaction to cessation of speech might be relevant given published minimal reaction times for spoken utterances (Fry, 1975; Izdebski and Shipp, 1978; Shipp et al., 1984). The distribution above this threshold represented 41–45% of all between-speaker intervals. These cases were thus potentially long enough to be reactions to the cessation of speech, or even more so to some prosodic information just before the silence."

There are two separate proposals here. The first is that for gaps longer than 200 ms, participants might simply react to silence. This threshold is implausible. First, silence will only become recognizable as silence after c. 200 ms (after all the duration of voiceless stop consonants ranges up to 180 ms; cf. Heldner and Edlund, 2010), at which point it will still take a further minimally 200 ms to react (so 400 ms in total). That minimal reaction is for a prepared vowel (Fry, 1975), and any more complex response will increase according to Hick's Law (see below); a choice between one of two prepared responses takes 350 ms for example. We now have, say, 550 ms from actual cessation of speech till beginning of a minimal response, and as Heldner and Edlund (2010) note 70–82% of responses are within 500 ms. Thus reaction to silence, although certainly possible in a minority of cases, would not seem to play a major role in the organization of turn-taking (see Riest et al., 2015).

The second proposal is that there is the possibility of reaction to "some prosodic information just before the silence." Here there is less room for disagreement; CA practitioners and associated phoneticians have themselves emphasized the role of turn-final intonational and segmental cues (see Walker, 2013 for a review). Duncan drew attention to turn-keeping intonation cues and lengthened ('drawled') syllables. Critical here are two factors: (a) it must be shown not only (as Duncan did) that there are available prosodic/phonetic features of turn-ends, but also that participants actually use them, (b) the location of the features with respect to the turn end is important (e.g., sentence accents in English sometimes occur well before turn ends, in which case talk of projection suits better than talk of reaction to terminal cues, cf. Wells and Macfarlane, 1998). Bögels and Torreira (in press) provide experimental evidence that listeners do use turn-final prosodic information (located in the last syllable of the utterance) to identify turn ends in Dutch questions with final rising intonation. Further research should investigate other linguistic contexts.

Another notion that has some currency is that turn-taking could be driven by coupled oscillators (Wilson and Wilson, 2005). Coupled oscillators have been shown to play a role in coordination in the animal world, e.g., in the synchronization of fire-fly flashing where an individual's flashes reset the neighboring fireflies' oscillators, so gradually converging on a single beat. However, it is well known that human synchronization does not primarily work in this way, but rather by means of temporal estimation, which is easily shown by demonstrating that humans can tap together without waiting to hear the others' taps (Buck and Buck, 1976). Moreover, given the highly variable lengths of turns, nothing like the firefly mechanism can work in conversation. Indeed, human coordination in general relies on simulating the other's task, thus on high-level cognition (Sebanz and Knoblich, 2008). There is, however, room for a low level metronome, as it were, and Wilson and Wilson (2005) suggest that readiness to speak is governed by the syllable, so that participant A's beginning of a syllable tends to coincide with B's least readiness to speak, while the end of the syllable coincides with B's increased readiness. There is indeed some evidence for entrainment or accommodation of the gap size between specific dyads, but there is no such effect on intra-turn pauses (ten Bosch et al., 2005) suggesting that turn-transition timing is rather unconnected to other temporal properties of speaking, although more research is required here.

Careful observers have convinced themselves that such a 'beat' is set up in English conversation by stress-timing, such that interlocutors producing unmarked actions with their turns tend to come in 'on the beat' (Couper-Kuhlen, 2009). However, the perceived rhythm of speech does not appear to have direct acoustic correlates, and to date we are unable to objectively confirm these observations (note too that languages differ in their rhythmic properties). Interestingly, recent corpus measurements show that, rather than the entrainment of a conversational beat, there is a reverse correlation of speaker A's speech rate and speaker B's response timing, perhaps because B has less time to plan her message as A's speech rate increases, and vice versa (Roberts et al., 2015).

5. Statistical Studies of Corpora

The statistical study of turn-taking began early, prompted by developments in telephony, with a special interest in the speed of turn-transition (e.g., Norwine and Murphy, 1938). It has become standard to represent overlaps and gaps on a single time scale [sometimes called 'the floor transfer offset' (FTO)] in which positive values correspond to gaps, and negative values represent overlap. **Table 1** summarizes average values of FTOs in ten languages as reported in four studies (caveat: codings and methods differ somewhat in these studies). Note that although mean values vary, they do so in narrow window, roughly a quarter of a second either side of the cross-linguistic mean, and that

TABLE 1 | Average floor transfer offsets (FTOs) in ten different languages as reported by four different studies.

Language	Average FTO (ms)	Source
English	410	Norwine and Murphy (1938)*
English	480	Sellen (1995)*
English	460	Sellen (1995)
Dutch	−78	De Ruiter et al. (2006)*
Japanese	7	Stivers et al. (2009)
Tzeltal	67	Stivers et al. (2009)
Yélî-Dnye	71	Stivers et al. (2009)
Dutch	108	Stivers et al. (2009)
Korean	182	Stivers et al. (2009)
English	236	Stivers et al. (2009)
Italian	309	Stivers et al. (2009)
Lao	419	Stivers et al. (2009)
Danish	468	Stivers et al. (2009)
ǂĀkhoe Hai‖ om	423	Stivers et al. (2009)

No eye-contact between conversation participants.

the factors affecting response times are uniform across cultures (Stivers et al., 2009). In the following two sections, we look in more detail at the distribution of gaps and overlaps.

5.1. Distribution of Gaps

About half a century ago, Brady (1968) reported average gap durations of 345–456 ms and medians from 264 to 347 ms (depending on the threshold used in the automatic detection of speech) in a corpus of sixteen telephone calls between friends in the USA. Task-oriented interaction shows surprisingly similar patterns [e.g., Verbmobil – a travel scheduling task by telephone, has geometric means of 380 ms (English), 363 ms (German), 389 ms (Japanese); Weilhammer and Rabold, 2003]. In a wide review, Heldner and Edlund (2010) looked at three different corpora, automatically processing two of them for speaker transitions: a Dutch dialog corpus, and English and Swedish Map Tasks (where interlocutors must adjust their positions on slightly mismatching maps). The first two corpora included both face-to-face and non-face-to-face interaction. Heldner and Edlund (2010) found closely matching patterns across corpora, with combined scale (FTO) modes for speaker transition at c. 200 ms (i.e., a short gap) and c. 60% of transitions being gaps, 40% overlaps (including any overlap of greater than 10 ms; the modal overlap is less than 50 ms in the Spoken Dutch Corpus). Around 41–45% of gaps were longer than 200 ms, and between 70 and 82% of all transitions were shorter than 500 ms.

These quantitative approaches generalize over all kinds of speech acts and responses. But there is also growing work focused specifically on question–answer timings. Question–answer sequences are an interesting context to examine, because questions make a floor transfer relevant, whereas in other contexts a floor transfer between speakers is often optional. Stivers et al. (2009) looked at 10 languages from around the world, including smaller, unwritten languages, and found rather fast transitions in polar question contexts, with means between 7 and 468 ms, and modes from 0 to 200 ms. The coding of this sample was from videotape and included early visual responses (e.g., nods) and audible pre-utterance inbreaths. The general finding was that although languages differ, e.g., in their degree of use of visual modality or mean response times, the factors that speeded or slowed response times (e.g., gaze, agreement) were shared. Heldner (2011) shows that estimates of the percentage of perceived overlaps and gaps in this sample match closely other quantitative samples.

The intensive study of turn-taking under different conditions is still in its infancy. We know that responses to Wh-questions are slower than polar (yes–no) questions cross-linguistically (unpublished data from the Stivers et al., 2009 study), presumably because of the greater cognitive complexity of response involved. Longer answers can also be shown to take more preparation, reflected in both reaction times, and breathing preparation (Torreira et al., 2015). Complexity of response has also been shown to influence timings in children's responses (Casillas, 2014). We also know that individuals tend to accommodate to the gap-length of others, so that when changing conversational partners, individuals' response times change to match their new interlocutors (ten Bosch et al., 2004, 2005). And intriguingly, transition speeds are higher on the phone than face-to-face (Levinson, 1983; ten Bosch et al., 2005).

5.2. Overlap

In contrast to gaps, the study of overlap in corpora has provided only gross facts. As mentioned, Heldner and Edlund (2010) report c. 40% of speaker-transitions involving overlaps (including any overlap of greater than 10 ms). Their histogram makes clear that the modal overlap is less than 50 ms in the Spoken Dutch Corpus, with a mean −610 ms, and median −470 ms. ten Bosch et al. (2005) report that the proportion of overlaps increases from 44% in face-to-face conversation to 52% in telephone conversation, with males more likely to overlap their interlocutor than females, but looking just at the transition from speaker A to speaker B, 80% of transitions are gaps and 20% partial overlaps in face-to-face conversation (the corresponding figures for telephony are 73 and 27%).

Because of the lack of detailed statistical analysis of overlaps in corpora, we have undertaken a new analysis of overlaps in the Switchboard Corpus of English telephone conversations (Godfrey et al., 1992). We address the following questions:

(1) In running speech, how common is overlap (i.e., simultaneous talk by more than one party at a time) compared to talk by one party alone?
(2) In floor transfers, how common are overlaps compared to gaps?
(3) What is the distribution of overlap duration, and where do overlaps tend to start relative to the interlocutor's turn?
(4) What is the distribution of different overlap types (cf. Jefferson, 1986)?

5.2.1. Method

We analyzed a subset of 348 conversations (totaling around 38 h of dyadic conversation) that were free of timing errors, and with annotations included in the NXT-Switchboard Corpus

release (Calhoun et al., 2010). To estimate the occurrence of overlaps in this dataset, we used the operationalization scheme in Heldner and Edlund (2010). First, based on the phonetic segmentation of the corpus, we divided each speaker's signal into interpausal units (IPUs) delimited by silent intervals of 180 ms or more. The 50,510 IPUs had an average duration of 1680 ms, and a median duration of 1227 ms. Second, we defined *gaps*, *pauses*, *between-overlaps,* and *within-overlaps* as follows. Gaps ($n = 14648$) corresponded to portions of the stereo signal that contained silence in each speaker's channel, and that involved a floor transfer between the two speakers. Between-overlaps ($n = 6524$) were floor transfers that occurred without a silent gap between the speakers, whereas within-overlaps ($n = 3343$) were parts of the signal with overlapping inter-pausal units that did not result in an effective floor transfer. **Figure 1** below illustrates the operationalization of gaps, between-overlaps and within-overlaps.

5.2.2. Findings

The recordings were divided as follows: 77% of the signal corresponded to speech by one speaker only, 19.2% to silence (i.e., either pauses within a speaker's turn or gaps as defined above), and only 3.8% to simultaneous speech by both speakers (either between-overlaps or within-overlaps). If we exclude silent parts, 95.3% of the speech signal corresponded to speech by one speaker. This seems to fit well with Sacks and colleagues' observation that "overwhelmingly, one party speaks at a time" (Sacks et al., 1974, p. 700).

With regard to how common overlaps are in terms of proportion of turn-transitions, **Figure 2** shows the distribution of the duration of gaps and between-overlaps combined together as FTOs (i.e., with positive values for gaps and negative values for between-overlaps). Between-overlaps (negative FTOs) represented 30.1% of all floor transfers. As for the duration of overlaps, and their location within the interlocutor's turn, we observed that between-overlaps exhibited a distribution highly skewed to the left, with an estimated modal duration of 96 ms, a median of 205 ms, a mean of 275 ms, and with 75% of the data with values below 374 ms. Within-overlaps tended to start close to the beginning of the utterances that they overlapped, with a modal offset of 350 ms, a median of 389 ms, a mean of 447 ms, and 75% of the data exhibiting offsets below 532 ms.

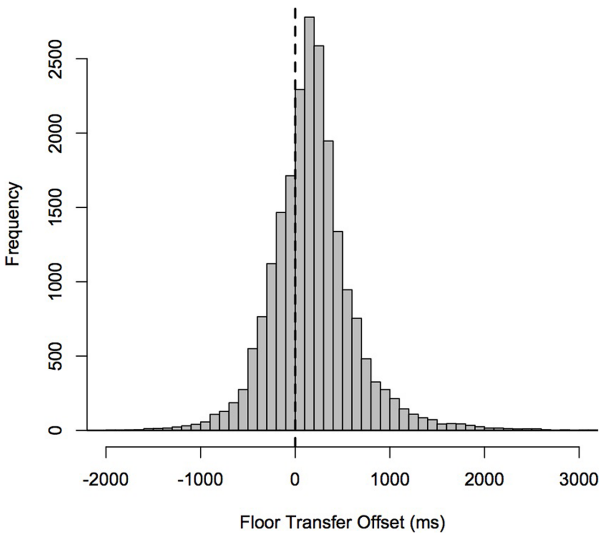


FIGURE 2 | Histogram of floor transfer offsets (FTOs) in the Switchboard Corpus (Godfrey et al., 1992; Calhoun et al., 2010, see Section 5.2.1 for details). Each bin has a size of 100 ms.

Their duration exhibited a distribution highly skewed to the right, with an estimated modal duration of 350 ms, a median of 389 ms, a mean of 447 ms, and 75% of the data with values below 532 ms. The duration of within-overlaps is thus usually shorter than that of two syllables. This appears to fit well with Sacks et al.'s (1974) observation that "occurrences of more than one speaker at a time are common, but brief."

We now examine the distribution of different types of overlaps. A prediction made by the Sacks et al. (1974) model is that most overlaps should be occasioned by a number of circumstances emerging from the application of its rules. For instance: (i) Overlaps often arise when unforeseen additions to the first speaker's turn after a transition relevance place (e.g., during increments or tags); (ii) They may occur after a silence when two speakers may self-select and launch articulation without realizing that another party is doing the same thing (cf. 'blind spot' cases, Jefferson, 1986); (iii) They may frequently arise in cases involving backchannels signaling feedback to the

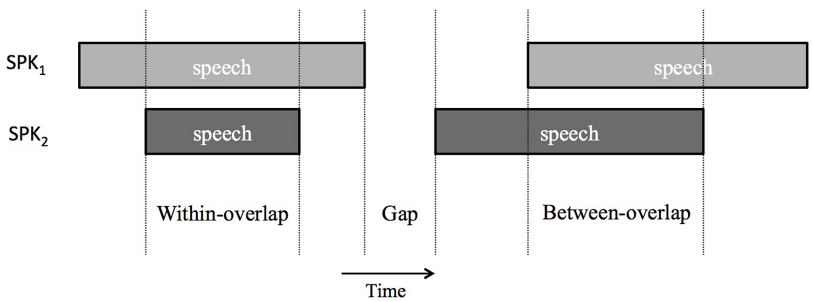


FIGURE 1 | Illustration of gaps, within-overlaps, and between-overlaps for two speakers (SPK_1 and SPK_2) in our classification scheme following Heldner and Edlund (2010).

main speaker (e.g., *yeah*, *right*) and other minimal utterances that do not constitute an attempt to take the floor. The Sacks et al. (1974) model also predicts signs of overlap avoidance when it occurs, for instance by speakers' abandoning their turns without reaching a point of turn completion. Another sign of speakers' special orientation to overlapping talk is that they may engage in competition for the floor, for instance by repeating syllables or words, often with increased intensity and pitch levels (Schegloff, 2000).

To estimate the prevalence of such possible causal contexts for overlap, in a separate analysis we randomly sampled 100 between-overlaps and 100 within-overlaps from our data, and annotated them for a number of relevant features, including (a) the presence of a backchannel or brief token of agreement (e.g., *yeah*, *right*) in either the overlapped or overlapping utterance, (b) the presence of a period of silence within 200 ms from the beginning of the overlap period, (c) the presence of a transition relevance place (a point of syntactic, prosodic and pragmatic completion) in the overlapped turn within the 500 ms leading to the overlap, (d) an abandoned (i.e., syntactically and prosodically incomplete) utterance by any of the two speakers during or immediately following the overlap interval, and (e) the presence of repeated syllables or words in any of the two speaker's utterances during or immediately following the overlap interval. Other recurrent features observed during or close to the overlap interval, such as laughter and disfluencies, were also annotated.

Table 2 shows the most frequent features observed in the data (note that the features are not mutually exclusive). Interestingly, the majority of overlap cases (73%) involved a backchannel. Backchannels, especially continuers like "mm hm" or "uh huh," are not construed as full turns, but rather pass up the opportunity to take a turn, and are thus principled intrusions into the other's speech (Schegloff, 2000). It should be noted that, in half of the between-overlaps, it was not the backchannel that incurred the overlap, but rather the main speaker who produced an utterance in overlap with the backchannel. We also noted that overlapping backchannels often occurred after a TRP or a period of silence, suggesting that their timing is sensitive to specific cues in the main speaker's turn (cf. Gravano and Hirschberg, 2009).

The second most common feature (37%) was the presence of a possible transition-relevance place (i.e., a point of syntactic, intonational, and pragmatic turn completion) in the overlapped turn within a time window of 500 ms before the start of the overlap. Another common feature was a period of silence (29%). In cases with this feature, one of the two speakers produced an utterance briefly after her interlocutor. These cases often involved a backchannel ($n = 35$, or 60%), or resulted in one of the two speakers abandoning their turn prematurely before reaching a point syntactic and prosodic completion ($n = 14$, or 24.1%). The presence of a disfluency in the utterance of the overlapped speaker before the start of the overlap (i.e., short silent pauses, repeated syllables or words, or noticeable decreases in speech rate) was also common. In these cases, it seems that the recipient produced a backchannel in response to the disfluency at a point when the interlocutor had already resumed her turn, causing overlap. In total, cases exhibiting one or more of these six features accounted for 95% of the data.

The remaining 10 cases involved three terminal between-overlaps affecting the last syllable of the previous turn, two cases exhibiting laughter by one of the speakers, two cases involving a turn-initial particle (i.e., *uhm* and *well*) produced in overlap with the last syllable of the preceding turn, one case with a speaker talking to someone else in the room, and one case of overlap due a clear phonetic segmentation error in the annotation.

Our analysis thus confirms that overlaps, though reasonably common (30% of transitions), are of short duration (i.e., less than 5% of the speech signal; between-overlaps have a modal duration 96 ms), occur largely in principled places (e.g., in between-overlaps, after possible completions, in simultaneous turn-starts), and mostly involve backchannels (which do not constitute full turns). In light of these observations, we conclude that the vast majority of instances of overlap in our dyadic conversations are consistent with the turn-taking system proposed by Sacks et al. (1974).

TABLE 2 | Frequency of seven features in a subset of 200 cases of overlap (100 between-overlaps, and 100 within-overlaps) extracted from our Switchboard data.

	Between-overlaps ($n = 100$)	Within-overlaps ($n = 100$)	Percentage in total ($n = 200$)
Backchannel or agreement present	74	72	73%
Follows TRP (<500 ms)	23	51	37%
Follows silence (simultaneous start)	21	37	29%
Abandoned turn	21	18	19.5%
Follows disfluency in interlocutor's turn	4	18	11%
Repeated syllables or words	4	12	8%
Any of the six features above	93	97	95%

Note that observations can exhibit more than one feature at the same time (e.g., cases of overlap after a period of silence involving a backchannel.

6. Psycholinguistics

Psycholinguistic processing puts tight constraints on any psychologically real model of turn-taking. Here we first draw attention to the early sensitivity to turn-taking in child development. Then we consider three main psycholinguistic aspects: predictive theories of language comprehension, studies of language production (from conceptual planning to speech articulation), and ideas about the relation between these two processes. Finally we turn to a small number of experimental studies aimed at understanding the relationship between comprehension and production processes in turn-taking.

6.1. 'Proto-Conversation' and Turn Taking in Human Development

Parallel to Sacks et al. (1974), in the 1970s there was an interest in children's acquisition of turn-taking abilities. Trevarthen

(1977) and Bruner (1983) coined the term "protoconversation" for the rhythmic alternation of vocalizations between caregiver and infant in the early months of life, and its systematic properties were demonstrated by Bateson (1975), with average turn transitions of about 1.5 s at 3 months. Subsequent work showed that this gap reduced in the following pre-linguistic months to around 800 ms (Jasnow and Feldstein, 1986; Beebe et al., 1988). Such early onset suggests that turn-taking may have an instinctive basis. Garvey and Berninger (1981) showed that the gap duration *increased* toward a second and a half in toddlers, presumably because of the cognitive difficulties of language production, and remained at around a second even for 5-year-olds [this slow convergence with adult norms has recently been confirmed for a larger sample by Stivers et al. (under review)].

After a long pause, there is now renewed interest in the development of turn-taking and its timing in children, and we now have better data, methods and concepts. Using audiovisual corpus techniques, Hilbrink et al. (submitted) have confirmed the general pattern earlier reported, namely relatively fast transitions in the prelinguistic period, with a slowing down as language starts to be comprehended at 9 months. Using eye-tracking of infants watching dyadic interaction, several studies have shown that 3-year-olds observers of dyadic conversations between two adults can *anticipate* speaker transitions (Tice and Henetz, 2011; Casillas and Frank, 2013, submitted; Keitel et al., 2013). Although the gaze shifts tend to occur in the gap (i.e., not in overlap with the turn preceding the floor transition), known saccade latencies for infants are c. 300 ms (Fernald et al., 2008), showing that they have often systematically detected the end of the turn before the gap. Researchers have also been able to show that by 3 year-olds, children are using intonation to do this projection of turn-ends (Keitel et al., 2013). Casillas and Frank (submitted) found that 3-year-olds were just as good at anticipating speaker change as adults, and did so more after questions than statements. They then looked at younger infants and filtered the speech, so they could distinguish whether prosody or lexico-syntax was enabling this anticipation. They found that 1 and 2 year-olds were better than chance at anticipating transitions, and that anticipation improves with age. Children under 3 were better in the prosody-only condition (with words filtered out) than they were in the words-only condition (with prosody filtered), indicating an early advantage for prosody (adults only showed an advantage for words + prosody). Clearly these studies confirm that projection is a real phenomenon, that it is learnt early, and that prosody plays an important role in this ability. They also indicate that turn-taking is established before language, that it forms a framework for language acquisition, and that the complexities of language slow down the framework through middle childhood.

6.2. Predictive Language Comprehension

Early in the history of psycholinguistics, Chomsky (1969, p. 57) insisted that probability and prediction had no possible role to play in a scientific theory of language: "It must be recognized that the notion 'probability of a sentence' is an entirely useless one, under any known interpretation of this term." He reasoned that a grammar bounds a discrete infinity, and hence there was no core role for prediction in language understanding. The spell lasted decades, but meanwhile both engineering and psycholinguistic experiments have demonstrated a core role for statistical learning and estimation in language comprehension. For example, eye-movement studies in the visual world paradigm show that listeners predict upcoming entities from likely collocations (e.g., hearing "the boy is eating" participants look at the cake and not the ball in the picture). Determiners (e.g., French *un* vs. *une*), Adjectives ("freshly baked") and verbs ("eat") can predict nouns by their selectional restrictions; in language that have verbs at the end of the sentence like Japanese, participants can use the nouns to predict the verbs (Altmann and Kamide, 1999; Kamide et al., 2003). Another source of insight comes from EEG, where it can be shown that the syntactic frame is used to predict upcoming material. For example, when the sentential context leads one to expect a specific noun ('she carried the eggs in a . . .') but the gender of an incoming article is incongruous an N400 may be evoked before the noun itself is encountered (e.g., in Spanish *una canasta* 'a basket' vs. *un costal* 'a sack'). These studies use the inverse correlation between the cloze probability and the amplitude of an N400 to demonstrate prediction (it is hard in fact to distinguish prediction from integration difficulties; see Kutas et al., 2011 for review). Predictive language comprehension is not only achieved on the basis of semantic and morphosyntactic regularities. In an experiment involving visual searches under the directions of a confederate, Ito and Speer (2008) showed that participants could anticipate referents on a screen (e.g., a "drum" vs. a "ball") on the basis of the location of contrastive pitch accents in the vocal instructions being given to them (e.g., "now take the GREEN ball" vs. "now take the green BALL"). Listeners therefore appear to be able to use different sorts of linguistic information (i.e., semantic, morphosyntactic, prosodic) in order to predict the content of an incoming utterance. For an overview of recent work on predictive language understanding see Pickering and Garrod (2013).

Recent investigations have also shown direct connections of these predictive inferences to projection in conversation. Gisladottir et al. (2015) conducted an EEG experiment in which participants listened to mini-dialogs of two turns. The second turn (e.g., "I have a credit card") could be invariant over three conditions, a question like "How are you going to pay?," an offer like "I can lend you the money," or a trouble announcement like "I don't have any money." In each of three contexts, the same second turn performs a different speech act (i.e., an answer, a declination, or an offer). The EEG signal, averaged over many such adjacency pairs, showed that very early (often in the first 400 ms) the different speech act forces of the response were predicted. Speech act detection is the precondition to response preparation, and it seems to be an early predictive process. A second relevant study (Magyari et al., 2014) looked at the EEG signal of participants listening to turns extracted from genuine conversations whose turn-endings they had to predict by pressing a button. These turns had already been sorted into unpredictable vs. predictable by a cloze test, where participants had to guess the missing words of items cut-off at various points. The predictable turns (compared to the unpredictable ones) showed a very early

EEG signature of preparation to respond about half way through the turn (c. 1200 ms before the end). Recently Riest et al. (2015) show experimentally that responses based on prediction are not significantly different than those based on pre-knowledge. They also incidentally attempt to estimate stochastic tendencies for possible reactive responses (although these stimuli are non-linguistic and do not have the uncertainty associated, e.g., with voiceless stops). These studies together suggest that quite long-range prediction is normally involved in understanding language in a conversational mode.

6.3. Latencies in Language Production

There are striking differences between language comprehension and production despite the fact that the processes must be intimately related. One of the clearest differences is in processing speed. Speech production is a bottleneck on the whole language system: at about an average of seven syllables per second, speech can be estimated to have a bit-rate of under 100 bps (Levinson, 2000, p. 28). Studies of language production show that pre-articulation processes run three or four times faster than actual articulation (Wheeldon and Levelt, 1995). Studies of language comprehension under compression show that people can parse and comprehend speech at three or four times the speed of speech production (Calvert, 1986, p. 178; Mehler et al., 1993). Speech encoding is one part of the process that has to be strictly serial. Articulation is thus a severe bottleneck on communication, and the system compensates by utilizing pragmatic heuristics in production that augment the coded message (Levinson, 2000).

Happily, there have been extensive studies of language production that allow us to quantify the latency in each part of the production process, using picture naming as a task (Levelt, 1989). The average reaction from seeing a picture to beginning the naming of has been estimated at 600 ms (Indefrey and Levelt, 2004, p. 106). The literature unfortunately gives no ranges or standard deviations, with the exception of a study by Bates et al. (2003), which provides cross-linguistic averages that are much longer at over 1000 ms, with all minimums over 650 ms. Indefrey and Levelt (2004, p. 108), on the basis of a meta-study of available experiments, propose approximate figures for each stage of the process, which we show in **Table 3**.

For multiword utterances, the effect is not linear. Naming two nouns takes 740–800 ms before output begins, with evidence that the processing of the second noun has begun but not finished by this time, while 900 ms is required for three word utterances (Schnur et al., 2006). Most of these studies incidentally (but not Bates et al., 2003) involve pre-familiarization of the words and pictures, so these response times are effectively after some amount of priming.

There is also good information on the planning required for sentence production from eye-movement studies. When participants are shown pictures of simple transitive or intransitive scenes (e.g., boy kicking ball, girl running), it takes about 1500 ms before speech output begins (Griffin and Bock, 2000; Gleitman et al., 2007). Interestingly, what happens within this 1500 ms is language-dependent – for example verb-first languages show rather different visual scanning of the pictures than verb medial languages (Norcliffe et al., 2015), but the latencies remain similar.

During this period of planning for language production, output processes involve the synergies between multiple speech organs. For example, breathing for speaking may need to be initiated. Earlier studies have shown that such breathing activity involves a number of latencies: first, c. 140–320 ms must be allowed for from the time the decision to inhale is made till the time the signal reaches the intercostal muscles (Draper et al., 1960); second, the inhalation time in spontaneous dialog is typically over 500 ms long (McFarland, 2001, p. 136). Together, these numbers suggest a latency of at least 500–800 ms prior to speech. In a recent study of breathing in conversation (Torreira et al., 2015, this volume), we have shown that short responses to questions are often made on residual lung air, whereas longer responses are likely to require a planned inhalation. The actual inhalation most typically starts briefly (i.e., 15 ms) after the end of the interlocutor's question, and it is probably triggered just before the phonological retrieval process for the first word of the planned response. Thus the breathing data suggests that whether or not inhalation is required is a decision made during conceptual planning of the response, and that the trigger for inhalation, most typically produced during the last few hundred ms of the interlocutor's turn, is often based on a prediction that the current speaker will imminently end her turn.

Recent studies of vocal preparation using ultrasound techniques show that tongue movements preceding speech production start considerably before the acoustic signal, with clear preparation between 120 and 180 ms prior to the acoustic release (Schaeffler et al., 2014) and with some effects detectable as early as 480 ms (Drake et al., 2014). Although not yet studied in a conversational context (although see de Vos et al., 2015, this volume, for the parallel in signed conversation), these measurements provide further estimates of the latencies involved in language production. These latencies are perhaps not surprising given the complexity of language encoding and the need for the processes to be funneled into a single, serial sequence of operations. Donders (1869) showed that reaction time varies with the number of choices that need to be made, and Hick's Law (Hick, 1952) suggests this relation is generally logarithmic (reaction time will increase with decision time, where decision time $T = \log_2(n)$ and n is the number of equally

TABLE 3 | Estimated average time windows for successive operations in spoken word encoding (Indefrey and Levelt, 2004, p. 108).

Operation	Duration (ms)
Conceptual preparation (from picture onset to selecting the target concept)	175
Lemma retrieval	75
Form encoding:	
Phonological code retrieval	80
Syllabification	125
Phonetic encoding (till initiation of articulation)	145
Total	600

probable choices). When one considers that in production single words have to be plucked from a word lexicon consisting of over 20,000 entries, one can see immediately the processing problems involved. Combined with the relatively slow nature of nerve conduction (known since Helmholtz, 1850), and the complexity of the coordination of c. 100 muscles involved in articulation (Levelt, 1989), slow reaction times can be expected.

To summarize, language production involves latencies of well over half a second, and a multi-word utterance is likely to involve a second or more of processing before articulation begins. Although the conversational context may expedite some of these processes, the bulk of this latency is attributed to the phonological and phonetic encoding processes (as are frequency effects, Jescheniak and Levelt, 1994) which are probably not compressible.

6.4. Experimental Studies of Turn-Taking

There have been as yet relatively few experimental studies of turn-taking, due to the difficulties involved in gaining sufficient experimental control in free interaction. However, indirect light has been thrown on the mechanisms by extracting turns from conversation and experimentally testing when and how participants detect turn ends. De Ruiter et al. (2006) extracted turns from a corpus of conversations in Dutch, and got participants to press a button in anticipation of turn endings. They manipulated the turns so that there were versions where pitch information was filtered out (No Pitch), where the words were masked but the pitch preserved (No Words), where both were filtered (No Pitch, No Words) and finally where amplitude variation was also removed (Noise condition). They found that accuracy of turn-end anticipation was preserved under No Pitch, but significantly lost under No Words, and hugely affected under the other conditions, and they claim that "The conclusion is clear: lexicosyntactic structure is necessary (and possibly sufficient) for accurate end-of-turn projection, while intonational structure, perhaps surprisingly, is neither necessary nor sufficient" (De Ruiter et al., 2006, p. 531).

This study suggested then that lexicon and syntax are the key guide to turn-structure and completion. But there are aspects of prosody and articulation that may be critical, and in the normal case intonation may also be an important signal. To test this, Bögels and Torreira (in press) used turns taken from multiple scripted interviews, with questions like "So you're a student at Radboud University?" (long version) vs. "So you're a student?" (short version). The short versions exhibited a higher maximum pitch and greater duration on the last syllable of the word 'student' than the long versions, due to the presence of an intonational phrase boundary at the end of this word in the short questions, but not in the long questions. They cross-spliced their materials in different ways, and did the same button-press experiment as De Ruiter et al. (2006). Participants often false alarmed (pressed the button) at 'student' when a phrase-final word was cross-spliced into the middle of the long version – they were clearly using the prosodic information to anticipate turn closure. Participants were also presented with truncated long sentences ending in a syntactic point of completion, but lacking a final intonation phrase boundary: now participants only reacted on average around 400 ms after the end of the stimulus, suggesting that in this case participants' button presses were produced in reaction to silence. On the other hand, in another condition consisting of similar words, but featuring a final intonational boundary, RTs were around 100 ms on average, suggesting reaction to or local prediction of an intonationally well-formed question end. It should be noted that while pitch had been filtered in the De Ruiter et al. (2006) study, duration and other phonetic cues to prosodic structure were still present in their filtered No Pitch condition. This new study shows that participants do use prosodic cues to judge turn-ending. What the de Ruiter et al. study does establish is that they need to be integrated with the lexical/syntactic information to carry turn-ending indications.

There are other experimental techniques that can be used to explore turn-taking. One is to use confederates (Bavelas and Gerwing, 2011), another to use the visual world paradigm with eye-tracking (Sjerps and Meyer, 2015). The latter study, using a dual task paradigm, found that maximal interference in the non-linguistic task occurred 500 ms before the end of the incoming turn (see also Boiteau et al., 2014); however, the linguistic task involved visual monitoring and was non-contingent with the incoming turn, so was far removed from conversation.

A method that combines control with live interaction involves alternating live and pre-recorded responses in such a way that participants are unaware of the manipulation (Bögels et al., 2014). In a recent study, we exploited this technique in a quiz-game (Bögels et al., submitted). Participants were recorded for EEG in a shielded room, and could not see the quiz master – this allowed some of the interaction to be live, some pre-recorded. The quiz questions were designed so that in some the answer was available early, and that in others the answer was available only toward the very end of the question, as in:

Which character, also called 007, appears in the famous movies? (Early)
Which character from the famous movies, is also called 007? (Late).

In a second experiment, participants heard the same questions but did not have to answer them. Instead, they only had to remember them, as prompted by later probes. The neural patterns were then compared with those in the first experiment, where participants had to verbally respond, to the second where they only had to comprehend and memorize. The results revealed a clear neural signature associated with production, localized in the appropriate areas, occurring within 500 ms of the point at which a plausible answer to the question became available. Bögels and colleagues interpreted this as showing that participants begin planning their response as soon as they can, up to a second or more before the incoming turn ends.

6.5. The Core Psycholinguistic Puzzle

From a psycholinguistic point of view, turn-taking presents the following puzzle: in spite of the long latencies involved in language production (600–1500 ms or more), participants often manage to achieve smooth turn transitions (with the most typical gaps as little as 100–300). As a solution to this puzzle, we suggest

that comprehension is predictive, even more so than is currently thought. As soon as possible, participants extract the speech act of the incoming utterance, which is the *sine qua non* for planning their appropriate response. In order to overcome the production latencies, they must also start the planning and encoding of the response as soon as possible.

This suggests that there is a significant overlap of comprehension and production processes. Given an average turn (approximated as an interpausal unit in our Switchboard Corpus data) of 1680 ms, somewhere in the middle response preparation may already be underway. This provides a second central puzzle: conversation involves constant double tasking, and this double tasking uses the same language system. The difficulty of the puzzle is increased when one takes into account the findings that both comprehension and production use much of the same neural circuitry (Segaert et al., 2011). It is plausible that the difficulty here is overcome through rapid task switching, and the gradual switch of resources from comprehension of the incoming turn toward production of the response.

Pickering and Garrod (2013) outline a general model of psycholinguistic processing, suggesting that production and comprehension are intimately intermeshed. Just as generally in action control, forward prediction of one's own action is performed to correct deviations, so in interaction forward prediction of the other's actions is used to check perception, and aid preparation of response. This is a nice account, but the complexities rapidly multiply. Listeners, on this account, are both using their full comprehension system, and running a fast simulation of the other's production in order to predict the outcome. Now, given the turn-taking facts established above, we must add to this computational burden the need to simultaneously prepare one's own turn in advance involving both the full production system and a hypothesized fast forward predictor. So the poor listener who is about to respond has not only the full comprehension and production processes running simultaneously, but also two fast prediction systems (one for self, one for other). This quadruple tasking looks unlikely, especially as similar tasks are hard to multitask. Additional problems are that unlike physical action prediction, which can be estimated by a few heuristics, it is not clear how a fast approximate language prediction system would be feasible especially in production – producers have to grind through the syntax to find, e.g., what order to put words in. More likely the real production system may be involved *minus* the phonological and phonetic encoding, which account for the bulk of the production latency.

In any case, regardless of how this is achieved, the experimental and corpus studies reviewed in this section converge in showing that participants in conversation often anticipate the content of the others' turns well in advance, and that they use that information to prepare their response early.

7. Models of Turn-taking

Let us now gather together how the observations and inferences discussed above constrain viable models of turn-taking. Any adequate model must be consistent with a number of observations and constraints, as originally noted by Sacks et al. (1974, p. 700). We are now, however, able to add both additional constraints and a certain amount of temporal precision to those early observations:

(1) Turns are mostly short (mean 1680 ms, median 1227 ms; cf. see Section 5.2.1), consisting of one or more interjections, phrases or clauses at the syntactic level, and one or more intonational units at the prosodic level. Turn ends typically co-occur with points of both syntactic and prosodic completion.
(2) Intra-speaker gaps are longer by c. 150 ms than inter-speaker gaps (ten Bosch et al., 2005), suggesting ordered rules (the rights to the next turn unit belong first to the next speaker, and only if not exercised, to the current speaker).
(3) Inter-speaker gaps are most typically short, with modal values for FTOs falling between 100 and 200 ms (cf. **Figure 2**). Medium gaps and short overlaps are also common, although less so than short gaps.
(4) Lengthy gaps (over 700 ms) may carry semiotic significance (mostly, of an undesired or unexpected response; Kendrick and Torreira, 2015), thus contributing to propel fast timing.
(5) Overlaps, though common, are brief (with a mean of 275 ms at turn-transitions, and occupying less than 5% of the spoken signal in our telephone calls data). Overlaps are more common at turn transitions than within turns, and mostly involve back-channels, simultaneous first-starts, disfluencies, and other features predicted by Sacks et al. (1974).
(6) Turn-taking is established early in infancy, long before full linguistic competence, which actually appears to slow down response times; adult conversation timing is not achieved till late in middle childhood.
(7) Given the latencies of speech production (over 600 ms), incoming turns have to be predicted if accurate timing is to be achieved. EEG recordings suggest the production process in responsive turns starts as soon as the gist of the incoming turn can be detected.
(8) Turn-final cues seem to be used to recognize that a turn is definitely coming to an end. These cues are typically prosodic (e.g., phrase-final syllable lengthening and specific melodic patterns in many intonational languages) but also syntactic (e.g., syntactic closure), and in principle could be of other types too (e.g., gestural). In the appropriate pragmatic context, these turn-final cues can trigger the decision of the next speaker to articulate. From the point of view of social interaction, it is effective articulation that constitutes a point of no return (as opposed to other preparatory events preceding speech, such as pre-utterance inhalations and mouth noises).

7.1. The Standard Model and Alternatives

We have outlined above the Sacks et al. (1974) model of turn-taking as an opportunity-based or sharing system, regulated by normative rules. The behavioral patterns on this account are the outcome of joint, coordinated determination of turns, against a background of an assumption of rights to minimal turns. Not all

turns are minimal of course, but in this case a bid must be made for an extended turn, as in:

(9) Terasaki, 1976, p. 53
D: I forgot to tell you the two best things that happen' to me today.
R: Oh super=What were they?
D: I got a B+ on my math test ((material omitted)) and I got an athletic award.

An alternative model is the turn-end signaling system proposed by Duncan (1972), also mentioned above, under which the system is wholly in the control of the current speaker, who has exclusive rights and signals transfer at the end of the turn. In contrast, Sacks et al. (1974) held that "It is misconceived to treat turns as units characterized by a division of labor in which the speaker determines units and boundaries," instead, "the turn as a unit is interactively determined."

Duncan (1972, p. 286) proposed a simple rule of the sort "The auditor may take his speaking turn when the speaker gives a turn-yielding signal." Such a system would be in effect like the "over and out" cuing at the end of turns on a two-way (half duplex) radio which permits hearing or talking but not both at once by a single party. Such a system predicts that overlap can only occur when "over" cues are mistakenly given or overridden; the large incidence of overlaps in corpora, and their clustering at principled locations (like overlapped tags or address forms) is then hard to reconcile with such a model. As mentioned, the model presumed that these turn-yielding signals such as intonation are context-independent, but in fact we know they are not – e.g., in English final rising intonation in a question may signal finality but in a statement continuation; thus their interpretation would have to be embedded in complex comprehension processes. The model is in any case very partial: it tells us nothing about how or why people should initiate a turn, why turns are generally short, how multiple participants can be integrated into a single conversation, how overlap is resolved, and so forth. But it may add a component to a more complex overall model.

7.2. Toward an Adequate Psycholinguistic Model of Turn Taking – Cognitive Processes in the Responder[1]

We believe that the property list in Section 7 above puts fairly narrow constraints on a possible model of turn-taking. One area of particular interest is the temporal constraints that turn-taking imposes on language processing, given that conversational interchange is the core form of language use. These constraints are funneled into one crucial link in the system, namely, the current addressee preparing to respond. Here we consider the cognitive processes that must be involved.

The crucial questions concern what factors govern the decision making process that lies behind the initiation and timing of response. While turn-final cues in the incoming turn seem likely to play a role, they cannot be sufficient given the long latencies in language planning and production. To overcome these long latencies, predictive comprehension must be involved, together with a strategy of early beginnings to production. Bögels et al. (submitted) suggest that production begins as soon as it can – that is, as soon as the speech act content of the incoming turn is clear. This implies of course dual-tasking, perhaps by rapid alternation ('time sharing'). A new study using a dual-task paradigm and eye-tracking suggests that the heaviest interference is rather late (Sjerps and Meyer, 2015), and tied to looking-for-speaking which was postponed in this task toward the end of the incoming turn. Both early and late processes are almost certainly involved, but what exactly is happening, and when during natural conversation remains to be determined.

The flowchart diagram in **Figure 3** sketches the cognitive processes that must minimally be at work in the recipient of a typical turn at talk during conversation. Predictive comprehension is underway early, and already by half way through more predictable turns will suggest a temporal envelope for completion (Magyari et al., 2014). If so, morphosyntax may provide most of the early clues to the overall structural envelope (e.g., turns beginning with *if* or *either* or *whenever* projects a two clause structure), so offering some long distance projection. Within the last half second or so, the actual words will often be predicted (Magyari and de Ruiter, 2012), and, within that same late time-frame, cues to imminent turn closure, usually prosodic and phonetic, are likely to appear (Local and Walker, 2012; Bögels and Torreira, in press), indicating a likely turn end.

A recipient's first task is to identify or predict the speech act or action being carried out – both the illocutionary force and the likely propositional content. In cases in which the illocutionary force of the incoming utterance makes a floor exchange relevant or due, production planning may begin as soon as it is recognized, as suggested by the results in Bögels et al. (submitted). Production is, at least in the latter stages, serial, and proceeds through conceptualization, lemma retrieval, phonological retrieval, and phonetic encoding, following a time course that seems well understood (Indefrey, 2011), extending 600–1200 ms or more before articulation depending on the ease of retrieval and the length of the turn. In this model, early preparation is assumed, but actual articulation is held till turn-final cues (e.g., upcoming syntactic closure, a non-turn-keeping intonational phrase boundary) are detected, whereupon actual articulation is launched. Assuming these cues fall in the last half-second of the incoming turn, reaction to those will be sufficient to launch pre-prepared material so that it appears soon after the other's turn is completed.

Figure 3 sketches the kind of interaction between comprehension and production processes that must be involved in a typical turn transition (i.e., involving a FTO of c. 200 ms). There is an early gist comprehension with speech act apprehension sent as soon as possible to the production conceptualizer (see Levinson, 2013; Gisladottir et al., 2015). The production system may automatically begin to formulate right down to the phonology (Bögels et al., submitted), but with the actual articulation held in a buffer until the comprehension

[1]The ideas presented in this section were developed in collaboration with Mathias Barthel, Sara Bögels, and the other members of the INTERACT project at the Max Planck Institute for Psycholinguistics. See also Section 5.3 in Heldner and Edlund (2010) for a parallel proposal.

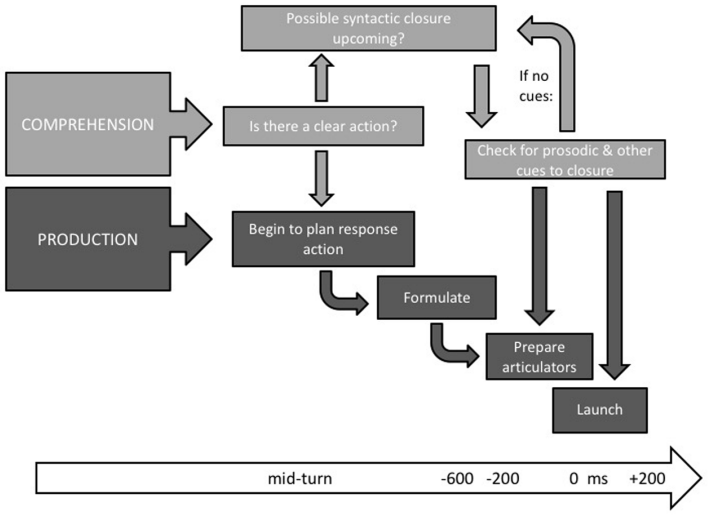


FIGURE 3 | Sketch of the interleaving of comprehension and production in the recipient of an incoming turn.

system signals an imminent completion of the incoming turn. Prior to that signal, it is likely that pre-articulation preparation (requiring c. 200 ms) of the vocal apparatus would be underway – this would include readying the vocal tract for the gestures to be made (see Drake et al., 2014; Schaeffler et al., 2014), and the decision to inhale prior to delivery of longer responses (Torreira et al., 2015, this volume).

Meanwhile the comprehension system continues to check the incoming signal for possible closure at both the syntactic and prosodic level. As soon as there are consistent signals of linguistic completion, a go-signal is sent to production, and any buffered articulation released. It is likely that visual monitoring of gesture can also be utilized for the go-signal (Duncan, 1974), but this awaits experimental confirmation.

This model is responsive to all the constraints listed in Section 7. What this model crucially adds is:

(a) an account of how responders can often respond with short latencies despite the long latencies of the production system;
(b) why the corpus statistical results reliably show a modal response with positive offsets of around 100–300 ms, reflecting the reaction time to the turn-final prosodic cues in the incoming turn (i.e., reaction to the go-signal, as hypothesized by Heldner and Edlund, 2010).

The model sketch in **Figure 3** is based on average, modal, and minimal temporal latencies reported in the literature. We would like to propose that this model is generally valid in the most frequent scenarios. If speakers launched their responses as early as they could without waiting for turn-final cues, we should expect overlapping or no-gap–no-overlap transitions to be the most common, rather than a short gap. And, if speakers typically launched language planning only after identifying turn-final cues, we should expect the most frequent transition times to involve at least half a second or more rather than short gaps of 100–300 ms. The model therefore captures the most typical turn transition values observed in conversational corpora.

What, however, accounts for the significant number of overlap and long gap cases observable in any conversation? A reviewer suggests that human factors such as lack of attention, pre-formulated agendas, and apparent involvement with actual minimal responsiveness may all be involved, and notes that apparent good timing may be achieved with buffers like particles. However, the evidence is that conversation is generally more demanding than that – for example 95% of questions get answers (Stivers, 2010), and particles like *well* and *uhm* in English are semiotically loaded and thus not empty buffers (Kendrick and Torreira, 2015), while Roberts et al. (2015) failed to find statistical differences in the timing of turns with and without such particles. In addition, it is likely that speakers sometimes use other turn-taking than the one sketched in **Figure 3**. For example, under competition for the floor, or when responding to highly predictable utterances, speakers may decide to launch articulation without waiting to identify turn-final cues. In cases of long transition latencies, speakers may not have been able to plan the initial stages of their turn early enough to launch articulation when the interlocutor's turn-final cues become available. This may indeed be due to a low attentional level on the part of the speaker, or to the interlocutor's turn being unclear in purpose until its end or simply to the complexity of the response required (Torreira et al., 2015, this volume).

8. Conclusion

This overview of work on turn-taking behavior over the last half century shows that turn-taking is a remarkable phenomenon, for it combines high temporal coordination between participants with the remarkable complexity and open-endedness of the language that fills the turns. The tension between these two properties is reflected in the development

of turn-taking in childhood (Proto-Conversation' and Turn Taking in Human Development), and it poses a substantial puzzle for psycholinguistic models (i.e., dual-tasking comprehension and production processes), which until recently have completely ignored this, the most central form of language use.

Acknowledgment

This research was financed by the ERC Advanced grant #269484 INTERACT and by the Max Planck Institute for Psycholinguistics.

References

Altmann, G., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychon. Bull. Rev.* 10, 344–380. doi: 10.3758/BF03196494

Bateson, M. C. (1975). Mother-infant exchanges: the epigenesis of conversational interaction. *Ann. N. Y. Acad. Sci.* 263, 101–113. doi: 10.1111/j.1749-6632.1975.tb41575.x

Bavelas, J. B., and Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *Int. J. Listening* 25, 178–198. doi: 10.1080/10904018.2010.508675

Beebe, B., Alson, D., Jaffe, J., Feldstein, S., and Crown, C. (1988). Vocal congruence in mother-infant play. *J. Psychol. Res.* 17, 245–259. doi: 10.1007/BF01686358

Bögels, S., Barr, D., Garrod, S., and Kessler, K. (2014). Conversational interaction in the scanner: mentalizing during language processing as revealed by MEG. *Cereb. Cortex* doi: 10.1093/cercor/bhu116 [Epub ahead of print].

Bögels, S., and Torreira, F. (in press). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *J. Phonet.*

Boiteau, T. W., Malone, P. S., Peters, S. A., and Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *J. Exp. Psychol. Gen.* 143, 295–311. doi: 10.1037/a0031858

Brady, P. T. (1968). A statistical analysis of on-off patterns in 16 conversations. *Bell Sys. Tech. J.* 47, 73–91. doi: 10.1002/j.1538-7305.1968.tb00031.x

Bruner, J. (1983). *Child's Talk*. New York, NY: Norton.

Buck, J., and Buck, E. (1976). 'Synchronous fireflies'. *Sci. Am.* 234, 74–85. doi: 10.1038/scientificamerican0576-74

Byrd, D. (1993). 54,000 American stops. *UCLA Work. Papers Phon.* 83, 97–116.

Calhoun, S., Carletta, J., and Brenier, J. M. (2010). The NXT-format switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Lang. Resour. Eval.* 44, 387–419. doi: 10.1007/s10579-010-9120-1

Calvert, D. R. (1986). *Descriptive Phonetics*, 2nd Edn. New York, NY: Thieme Medical Publisher.

Casillas, M. (2014). "Taking the floor on time: delay and deferral in children's turn taking," in *Language in Interaction: Studies in Honor of Eve V. Clark*, eds I. Arnon, M. Casillas, C. Kurumada, and B. Estigarribia (Amsterdam: Benjamins), 101–114. doi: 10.1075/tilar.12.09cas

Casillas, M., and Frank, M. C. (2013). "The development of predictive processes in children's discourse understanding," in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Society), 299–304.

Clayman, S. (2013). "Turn-constructional units and the transition-relevance place," in *Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Chichester: Wiley-Blackwell), 151–166.

Chomsky, N. (1969). "Quine's Empirical Assumptions," in *Words and Objections*, eds D. Davidson and J. Hintikka (Dordrecht: Reidel), 53–68. doi: 10.1007/978-94-010-1709-1_5

Couper-Kuhlen, E. (2009). "Relatedness and timing in talk-in-interaction," in *Where Prosody Meets Pragmatics*, eds D. Barth-Weingarten, N. Dehé, and A. Wichmann (Leiden: Brill), 257–276. doi: 10.1163/9789004253223_012

Crystal, T., and House, A. (1988). Segmental durations in connected-speech signals: current results. *J. Acoust. Soc. Am.* 83, 1553–1573. doi: 10.1121/1.395911

De Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

de Vos, C., Torreira, F., and Levinson, S. C. (2015). Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. *Front. Psychol.* 6:268. doi: 10.3389/fpsyg.2015.00268

Donders, F. C. (1869). "On the speed of mental processes," in *Attention & Performance II*, ed. and trans. W. G. Koster (Amsterdam: North-Holland), 412–431.

Drake, E., Schaeffler, S., and Corley, M. (2014). "Articulatory effects of prediction during comprehension: an ultrasound tongue imaging approach," in *Proceedings of the 10th International Seminar on Speech Production*, Cologne.

Draper, M. H., Ladefoged, P., and Whitteridge, D. (1960). Expiratory pressures and air flow during speech. *Br. Med. J.* 1, 1837–1843. doi: 10.1136/bmj.1.5189.1837

Drew, P. (2013). "Turn Design," in *Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Chichester: Wiley-Blackwell), 131–149.

Duncan, S. D. (1972). Some signals and rules for taking speaking turns in conversation. *J. Pers. Soc. Psychol.* 23, 283–292. doi: 10.1037/h0033031

Duncan, S. D. (1974). On the structure of speaker-auditor interaction during speaking turns. *Lang. Soc.* 2, 161–180. doi: 10.1017/S0047404500004322

Fernald, A., Zangl, R., Portillo, A. L., and Marchman, V. A. (2008). "Looking while listening: using eye movements to monitor spoken language comprehension by infants and young children," in *Developmental Psycholinguistics: On-line Methods in Children's Language Processing*, eds I. A. Sekerina, E. M. Fernandez, and H. Clahsen (Amsterdam: Benjamins), 97–135. doi: 10.1075/lald.44.06fer

Ford, C. E., and Thompson, S. A. (1996). "Interactional units in conversation: syntactic, intonational, and pragmatic resources for the projection of turn completion," in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 135–184.

Fry, D. B. (1975). Simple reaction-times to speech and non-speech stimuli. *Cortex* 11, 355–360. doi: 10.1016/S0010-9452(75)80027-X

Garvey, C., and Berninger, G. (1981). Timing and turn-taking in children's conversations. *Discourse Process.* 4, 27–57. doi: 10.1080/01638538109544505

Gisladottir, R., Chwilla, D., and Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLoS ONE* 10:e0120068. doi: 10.1371/journal.pone.0120068

Gleitman, L. R., January, D., Nappa, R., and Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *J. Mem. Lang.* 57, 544–596. doi: 10.1016/j.jml.2007.01.007

Godfrey, J., Holliman, E., and McDaniel, J. (1992). "SWITCHBOARD: telephone speech corpus for research and development," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (San Francisco, CA: IEEE), 517–520. doi: 10.1109/icassp.1992.225858

Goodwin, C. (1980). Restarts, pauses, and the achievement of mutual gaze at turn-beginning. *Soc. Inq.* 50, 272–302. doi: 10.1111/j.1475-682X.1980.tb00023.x

Gravano, A., and Hirschberg, J. (2009). "Backchannel-inviting cues in task-oriented dialogue," in *Proceedings of SigDial 2009*, London, 253–261.

Griffin, Z. M., and Bock, K. (2000). What the eyes say about speaking. *Psychol. Sci.* 4, 274–279. doi: 10.1111/1467-9280.00255

Hayashi, M. (2013). "Turn allocation and turn sharing," in *Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Chichester: Wiley-Blackwell), 167–190.

Heldner, M. (2011). Detection thresholds for gaps, overlaps and no-gap-no-overlaps. *J. Acoust. Soc. Am.* 130, 508–513. doi: 10.1121/1.3598457

Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *J. Phon.* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002

Helmholtz, H. (1850). "Vorläufiger Bericht Über die Fortpflanzungs-Geschwindigkeit der Nervenreizung," in *Archiv für Anatomie, Physiologie und wissenschaftliche Medicin* (Berlin: Veit & Comp.), 71–73.

Hick, W. E. (1952). On the rate of gain of information. *Q. J. Exp. Psychol.* 4, 11–26. doi: 10.1080/17470215208416600

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2:255. doi: 10.3389/fpsyg.2011.00255

Indefrey, P., and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001

Ito, K., and Speer, S. R. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *J. Mem. Lang.* 58, 541–573. doi: 10.1016/j.jml.2007.06.013

Izdebski, K., and Shipp, T. (1978). Minimal reaction times for phonatory initiation. *J. Speech Hear. Res.* 21, 638–651. doi: 10.1044/jshr.2104.638

Jasnow, M., and Feldstein, S. (1986). Adult-like temporal characteristics of mother-infant vocal interactions. *Child Dev.* 57, 754–761. doi: 10.2307/1130352

Jefferson, G. (1984). "Notes on some orderliness of overlap onset," in *Discourse Analysis and Natural Rhetoric*, eds V. D'Urso and P. Leonardi (Padua: Cleup Editore), 11–38.

Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Hum. Stud.* 9, 153–183. doi: 10.1007/BF00148125

Jescheniak, J. D., and Levelt, W. J. M. (1994). Word frequency effects in speech production: retrieval of syntactic information and of phonological form. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 824–843. doi: 10.1037/0278-7393.20.4.824

Kamide, Y., Altmann, G. T. M., and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J. Mem. Lang.* 49, 133–156. doi: 10.1016/S0749-596X(03)00023-8

Keitel, A., Prinz, W., Friederici, A. D., von Hofsten, C., and Daum, M. M. (2013). Perception of conversations: the importance of semantics and intonation in children's development. *J. Exp. Child Psychol.* 116, 264–277. doi: 10.1016/j.jecp.2013.06.005

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol.* 26, 22–63. doi: 10.1016/0001-6918(67)90005-4

Kendrick, K., and Torreira, F. (2015). The timing and construction of preference: a quantitative study. *Discourse Process.* 52, 255–289. doi: 10.1080/0163853X.2014.955997

Kutas, M., DeLong, K. A., and Smith, N. J. (2011). "A look around at what lies ahead: prediction and predictability in language processing," in *Predictions in the Brain: Using our Past to Generate a Future*, ed. M. Bar (Oxford: Oxford University Press), 190–207.

Lerner, G. H. (1991). On the syntax of sentences in progress. *Lang. Soc.* 20, 441–458. doi: 10.1017/S0047404500016572

Lerner, G. H. (2002). "Turn-sharing: the choral co-production of talk-in-interaction," in *The Language of Turn and Sequence*, eds C. Ford, B. Fox, and S. Thompson (Oxford: Oxford University Press), 225–256.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. (2000). *Presumptive Meanings*. Cambridge, MA: MIT Press.

Levinson, S. (2013). Recursion in pragmatics. *Language* 89, 149–162. doi: 10.1353/lan.2013.0005

Local, J., and Walker, G. (2012). How phonetic features project more talk. *J. Int. Phon. Assoc.* 42, 255–280. doi: 10.1017/S0025100312000187

Magyari, L., Bastiaansen, M. C. M., De Ruiter, J. P., and Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *J. Cogn. Neurosci.* 26, 2530–2539. doi: 10.1162/jocn_a_00673

Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *J. Speech Lang. Hear Res.* 44, 128–143. doi: 10.1044/1092-4388(2001/012)

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., and Pennebaker, J. W. (2007). Are women really more talkative than men? *Science* 317, 82. doi: 10.1126/science.1139940

Mehler, J., Sebastian, N., Altmann, G., Christophe, A., and Pallier, C. (1993). Understanding compressed sentences: the role of rhythm and meaning. Paper presented at the Temporal information processing in the nervous system. *Ann. N. Y. Acad. Sci.* 682, 272–282. doi: 10.1111/j.1749-6632.1993.tb22975.x

Munhall, K., Gribble, P., Sacco, L., and Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept. Psychophys.* 58, 351–362. doi: 10.3758/BF03206811

Norcliffe, E., Konopka, A., Brown, P., and Levinson, S. C. (2015). Word order affects the time-course of sentence formulation in Tzeltal. *Lang. Cogn. Neurosci.* doi: 10.1080/23273798.2015.1006238

Norwine, A. C., and Murphy, O. J. (1938). Characteristic time intervals in telephonic conversation. *Bell Syst. Tech. J.* 17, 281–291. doi: 10.1002/j.1538-7305.1938.tb00432.x

Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495

Pomerantz, A., and Heritage, J. (2013). "Preference," in *Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Chichester: Wiley-Blackwell), 210–228.

Riest, C., Jorschick, A. B., and De Ruiter, J. P. (2015). Anticipation in turn-taking: mechanisms and information sources. *Front. Psychol.* 6:89. doi: 10.3389/fpsyg.2015.00089

Roberts, F., Margutti, P., and Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Process.* 48, 331–354. doi: 10.1080/0163853X.2011.558002

Roberts, S. G., Torreira, F., and Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Front. Psychol.* 6:509. doi: 10.3389/fpsyg.2015.00509

Rossano, F. (2013). "Gaze in conversation," in *Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Chichester: Wiley-Blackwell), 308–329.

Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Schaeffler, S., Scobbie, J. M., and Schaeffler, F. (2014). "Measuring reaction times: vocalisation vs. *articulation*," in *Proceedings of the 10th International Seminar on Speech Production*, Cologne.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29, 1–63. doi: 10.1017/S0047404500001019

Schnur, T. T., Costa, A., and Caramazza, A. (2006). Planning at the phonological level during sentence production. *J. Psycholinguist. Res.* 35, 189–213. doi: 10.1007/s10936-005-9011-6

Sebanz, N., and Knoblich, G. K. (2008). "From mirroring to joint action," in *Embodied Communication*, eds I. Wachsmuth, M. Lenzen, and G. K. Knoblich (Oxford: Oxford University Press), 129–150.

Segaert, K., Menenti, L., Weber, K., and Hagoort, P. (2011). A paradox of syntactic priming: why response tendencies show priming for passives, and response latencies show priming for actives. *PLoS ONE* 6:e24209. doi: 10.1371/journal.pone.0024209

Sellen, A. J. (1995). Remote conversations: the effects of mediating talk with technology. *Hum. Comput. Interact.* 10, 401–444. doi: 10.1207/s15327051hci1004_2

Shipp, T., Izdebski, K., and Morrissey, P. (1984). Physiologic stages of vocal reaction time. *J. Speech Hear. Res.* 27, 173–178. doi: 10.1044/jshr.2702.173

Sjerps, M., and Meyer, A. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition* 136, 304–324. doi: 10.1016/j.cognition.2014.10.008

Stivers, T. (2010). An overview of the question-response system in American English conversation. *J. Pragmatics* 42, 2772–2781. doi: 10.1016/j.pragma.2010.04.011

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Stivers, T., Enfield, N. J., and Levinson, S. C. (2010). Question-response sequences in conversation across ten languages: an introduction. *J. Pragmatics* 42, 2615–2619. doi: 10.1016/j.pragma.2010.04.001

ten Bosch, L., Oostdijk, N., and Boves, L. (2005). On temporal aspects of turn-taking in conversational dialogues. *Speech Commun.* 47, 80–86. doi: 10.1016/j.specom.2005.05.009

ten Bosch, L., Oostdijk, N., and de Ruiter, J. P. (2004). "Turn-taking in social talk dialogues: temporal, formal, and functional aspects," in *Proceedings of the Ninth Conference on Speech and Computer* (*SPECOM 2004*), Saint-Petersburg: St.Petersburg, 454–461.

Terasaki, A. (1976). *Pre-announcement Sequences in Conversation* (*No. 99*). Irvine, CA: University of Irvine, Social Sciences.

Tice, M., and Henetz, T. (2011). "Turn-boundary projection: looking ahead," in *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, eds L. Carlson, C. Hölscher, and T. Shipley (Austin, TX: Cognitive Science Society), 838–843.

Torreira, F., Bögels, S., and Levinson, S. C. (2015). Breathing for answering: the time course of response planning in conversation. *Front. Psychol.* 6:284. doi: 10.3389/fpsyg.2015.00284

Trevarthen, C. (1977). "Descriptive analyses of infant communicative behaviour," in *Studies in Mother-Infant Interaction*, ed. H. R. Schaffer (London: Academic Press), 89–117.

Walker, G. (2013). "Phonetics and prosody in conversation," in *Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Chichester: Wiley-Blackwell), 455–474.

Weilhammer, K., and Rabold, S. (2003). "Durational aspects in turn taking," in *Proceedings of the International Conference of Phonetic Sciences*, Barcelona.

Wells, B., and Macfarlane, S. (1998). Prosody as an interactional resource: turn-projection and overlap. *Lang. Speech* 41, 265–294. doi: 10.1177/002383099804100403

Wheeldon, L. R., and Levelt, W. J. M. (1995). Monitoring the time-course of phonological encoding. *J. Mem. Lang.* 34, 311–334. doi: 10.1006/jmla.1995.1014

Wilson, M., and Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* 12, 957–968. doi: 10.3758/BF03206432

Wilson, T. P., and Zimmerman, D. H. (1986). The structure of silence between turns in two-party conversation. *Discourse Process.* 9, 375–390. doi: 10.1080/01638538609544649

Yngve, V. H. (1970). On getting a word in edgewise. *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 11 June 2015
doi: 10.3389/fpsyg.2015.00751

The use of content and timing to predict turn transitions

Simon Garrod[1]* *and Martin J. Pickering*[2]

[1] *Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK,* [2] *Department of Psychology, University of Edinburgh, Edinburgh, UK*

For addressees to respond in a timely fashion, they cannot simply process the speaker's utterance as it occurs and wait till it finishes. Instead, they predict both when the speaker will conclude and what linguistic forms will be used. While doing this, they must also prepare their own response. To explain this, we draw on the account proposed by Pickering and Garrod (2013a), in which addressees covertly imitate the speaker's utterance and use this to determine the intention that underlies their upcoming utterance. They use this intention to predict when and how the utterance will end, and also to drive their own production mechanisms for preparing their response. Following Arnal and Giraud (2012), we distinguish between mechanisms that predict timing and content. In particular, we propose that the timing mechanism relies on entrainment of low-frequency oscillations between speech envelope and brain. This constrains the context that feeds into the determination of the speaker's intention and hence the timing and form of the upcoming utterance. This approach typically leads to well-timed contributions, but also provides a mechanism for resolving conflicts, for example when there is unintended speaker overlap.

Keywords: dialog, turn-taking, prediction, timing, content

OPEN ACCESS

Edited by:
Judith Holler, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Nai Ding, New York University, USA
Alex Fine, Hebrew University, Israel

***Correspondence:**
Simon Garrod, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8YR, UK
simon.garrod@glasgow.ac.uk

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *29 October 2014*
Accepted: *20 May 2015*
Published: *11 June 2015*

Citation:
Garrod S and Pickering MJ (2015) The use of content and timing to predict turn transitions. Front. Psychol. 6:751. doi: 10.3389/fpsyg.2015.00751

Introduction

How is it possible for most conversations to be so fluent and efficient? Interlocutors tend to respond coherently and appropriately to each other. But in addition, they do so in good time—they do not leave long gaps between contributions, nor do they speak simultaneously for more than a brief moment (Sacks et al., 1974). To understand this remarkable and almost universal ability for turn transitions, we need to explain the cognitive processes that take place in people's minds. So far, psychologists have developed detailed accounts of the moment-by-moment processes that underlie producing and comprehending in isolation, but have much less to say about the moment-by-moment processes involved in conversation. In this paper, we propose an account of those processes that specifically explains turn transitions.

The Nature of Turn Transitions

We begin with an example from Schegloff (1996, p. 73). Two students are talking on the phone about a book purchase, with [indicating overlapping speech, and numbers indicating noticeable pauses in seconds)[1,2].

1.
Bee: I'nna tell you on:e course.
(0.5)
Ava: [( ).]
Bee: [(The mah-] the mah:dern art. The twunnieth century a:rt
there's about eight books,
Ava: Mm[hm,
Bee: [En I wen tuh buy a book the other day I [went] 'bh went=
Ava: [(mm)]
Bee: =down tuh N.Y.U. tuh get it becuz it's the only place thet
car[ries the book
Ava: [Mmm
Ava: Mmh
Bee: Tch! En it wz twun::ty do::lliz.
Ava: Oh my god.
(0.4)
Bee: Yeuh he-ez he wz handing me the book en 'etol' me twunny
dolliz I almos' dro(h)pped i(h)[t 'hh 'hh
Ava: [thhunh.
Bee: 'hhh I said but fer twunny dollars I bettuh hh 'hh yihknow,
(0.2)
Bee: 'hhh h[hold o:nto i(h)hh] huhh huh] 'hh!
Ava: [not drop it.] huhh huh]
(0.2)
Bee: Ih wz, (0.2) y'know (fun).=...

It is quite clear that the interlocutors contribute sequentially. On the one hand, any pauses are very short, but on the other, there is little overlap. In most cases, the overlap is not likely to interfere with comprehension, because people are able to speak and comprehend "backchannel" contributions such as *mm* (or listen and produce such contributions) at the same time. Somehow, the addressee must be able to know when to speak and when to be quiet, even though she does not know exactly what her partner is going to say.

Conversation analysts have very carefully analyzed what people do during conversations such as these (i.e., rather than highly ritualized or formulaic interchanges). Among other observations, Sacks et al. (1974, pp. 700–701) noted:

(a) Overwhelmingly, one party talks at a time.
(b) Occurrences of more than one speaker at a time are common, but brief.
(c) Transitions (from one turn to the next) with no gap and no overlap are common. Together with transitions characterized by slight gap or slight overlap, they make up the vast majority of transitions.
(d) Turn size is not fixed, but varies.
(e) What parties say is not specified in advance.
(f) Turn-allocation techniques are obviously used. A current speaker may select a next speaker (as when he addresses a question to another party); or parties may self-select in starting to talk.
(g) Repair mechanisms exist for dealing with turn-taking errors and violations; e.g., if two parties find themselves talking at the same time, one of them will stop prematurely, thus, repairing the trouble.

All of these observations are clearly relevant for the above example. Our goal is to explain such observations in cognitive-psychological terms. Our focus is on (c), and to some extent (b), (f), and (g). One important reason for (a) is presumably basic limitations on processing resources (it is very hard to produce and comprehend different messages at the same time). Observations (d) and (e) occur because conversations are generally unplanned and because people's goals vary (they may want to make small or large contributions) and may be affected by the conversation itself.

Sacks et al.'s (1974) work is based on English. Stivers et al. (2009) compared turn transitions for questions and responses across speakers of 10 diverse languages and found slight variation in distribution. But in all cases the most frequent interval was between 0 and 200 ms. In other words, conversationalists show a strong disposition to avoid overlap and to minimize silence between turns. They concluded that these properties of conversation constitute robust human universals (though cultural and linguistic factors lead to minor variations). So how is it possible for interlocutors to contribute with such short intervals between turns, while avoiding extensive overlap? How can the addressee prepare and execute an appropriate response while comprehending the speaker?

The Processes Underlying Turn Transition

Given such intervals, addressees cannot simply wait for the speaker to end before preparing their response. First, it would of course take some time to determine that the speaker has ended.

[1]Transcription conventions for the original examples in our paper are as follows: (.) indicates brief pause; [indicates overlap; : indicates lengthening; *.hh* indicates long inhalation; (=) at the end of one line and the beginning of the next indicate that the speech is continuous; ( ) indicates that the speech here was unintelligible to the transcriber. Speech between >XXX< is more compressed in pace, speech between <XXX> is more stretched out in pace. Upper case indicates unusually loud speech. In some cases we have added bold font for illustrative purposes.

[2]A more literal rendition of this exchange is:
Bee: I'm going to tell you about one course. The modern art – the twentieth century art – there's about eight books.
Ava: mm
Bee: And I went to buy a book the other day –
Ava: mm
Bee: went down-town to NYU to get it because it's the only place that carries the book.
Ava: mm
Bee: And it was twenty dollars.
Ava: Oh my god.
Bee: Yeah he was handing me the book and he told me twenty dollars. I almost dropped it -
Ava: huh
Bee: I said for twenty dollars I better you know-
Bee: hold onto it
Ava: Not drop it
Bee: I was - you know....

Second, many studies have demonstrated that producing a single word requires about 175 ms to access meaning, 75 ms for syntax, 205 ms for phonology, and 145 ms for phonetic encoding and articulation (Indefrey and Levelt, 2004; see Sahin et al., 2009, for somewhat different estimates). Even if these timings might be slightly different in conversation (rather than, for example, picture naming), it is clear that, in general, addressees must be able to estimate when the speaker's turn will end and begin response preparation several hundred milliseconds before that point.

This suggests that comprehension and production processes must be tightly interwoven. In fact, this assumption is quite controversial within the psychology of language, which tends to have studied comprehension and production in isolation and assumes that they involve largely independent mechanisms (Pickering and Garrod, 2013a). According to traditional accounts, dialog therefore can be characterized as serial monolog, in which the speaker produces and the addressee listens, and at the turn-transition point (i.e., transition relevance place) they switch roles and processes.

In fact, the serial monolog account suggests that speakers cannot prepare their utterances until they realize that their partner has completed (which may be later than the actual completion point). This would obviously be incompatible with Sacks et al. (1974) and Stivers et al. (2009). To avoid these problems, comprehenders would have to use ancillary mechanisms based on their comprehension systems to predict turn completions. These mechanisms would not be relevant for production, so they would have to begin preparing a response using their production systems in parallel to comprehension-based prediction. Moreover, they would have to determine the meaning of the complete utterance and then use this as a basis for generating an appropriate response.

These problems are, however, avoided if comprehenders use their production systems to make predictions and prepare their responses together. The mechanisms that they use to predict a speaker's final word, for example, are closely related to the mechanisms they themselves use to produce their response—or indeed to complete their partner's utterance if necessary (e.g., to help with word finding difficulty; A: *That tree has, uh, uh* … B: *tentworms*; Clark and Wilkes-Gibbs, 1986, p. 6). We now (1) specify the problem faced by the addressee; (2) discuss how addressees use predictions of timing and content to predict when the speaker will complete; and (3) discuss how the addressee can produce an appropriate and timely response.

Managing Fluent Turn-Transition Requires Predicting Both Speech Content and Timing

The addressee has to predict when the speaker is going to finish, and prepare an appropriate response. It would not be sufficient to prepare a response without predicting the end-point, because studies have shown that producing a prepared linguistic response to a cue takes several hundred milliseconds. For example, Ferreira (1991) had participants memorize and then produce sentences following a cue, and found response times of 500 ms or more. It takes at least as long to initiate prepared picture naming (e.g., Piai et al., 2011). Similarly, simply predicting the end-point would merely remove any time needed to determine that the speaker had ended, but not help with response preparation.

In fact, De Ruiter et al. (2006) showed that listeners could accurately estimate when a speaker's conversational turn was about to end. Their participants heard turns taken from recordings of natural conversations and indicated precisely when they thought the turn would end. The average response was about 186 ms before the turn actually ended. Interestingly, turn-ending estimates were not affected by flattening the pitch contour of the speech, but were dramatically affected when the lexico/semantic content was removed. This suggests that listeners used the content to predict turn endings. It is of course possible that other sources of prosodic information might affect estimates; for example, future investigations could test whether addressees are sensitive to rising intonation when responding to a question. In a subsequent study, Magyari and De Ruiter (2012) had another group of listeners predict the remaining words in De Ruiter et al.'s turn fragments. They found turn-end judgments were more precise when those listeners made accurate predictions than when they did not. An obvious explanation of these findings is that people's predictions of words constitute a factor (alongside speech rate) that is used to predict turn endings.

Experimental studies have shown that people predict aspects of upcoming words such as their syntactic features (e.g., Van Berkum et al., 2005) and their sound (e.g., DeLong et al., 2005), and that they also predict upcoming constituent structure (Staub and Clifton, 2006). Indeed, many theoretical accounts assume that comprehension is an inherently predictive process (Hale, 2001; Levy, 2008). We therefore propose that people can draw on local predictions of words and other linguistic information to predict turn endings.

It may also be possible to make predictions relating to semantics and pragmatics over a much longer period. The semantics of the context will place great constraints on the upcoming content (e.g., whether the speaker is likely to talk about food, work, holiday plans, or whatever). Of course, such information can come from the current utterance (e.g., *Changing the subject, I'm hungry—what would you* …). Sometimes this information will only be apparent just before the prediction is needed, but often the relevant words occur early in the utterance, or in a previous utterance. In other cases, the information comes from the non-linguistic context (e.g., an unfolding event such as a parade), or from shared background knowledge (i.e., common ground). Usually, this information is available well before the prediction is needed. The addressee also benefits from determining the speaker's speech act before it is complete, because whether the speaker is producing a statement, question, or command may help determine the upcoming length of the utterance. (As we discuss later, determining the speech act is also critical to preparing a response).

From these sources of information, the addressee could predict what the speaker is likely to say. These predictions could

include determining how much the speaker has left to say, as well as what the speaker is talking about. But to determine when the speaker will finish, the addressee has to combine these predictions with information about the speaker's speech rate. As our focus is on turn-transition, we now consider prediction of remaining content (what the speaker has left to say) and precise timing (when the speaker will finish). We then show how these predictions feed into the content and timing of the response itself.

Using Prediction-by-Simulation in Turn Transition

Now we account for addressees' ability to predict turn-ending and deliver an appropriate and timely response. To do this, we draw on the integrated account of production and comprehension developed by Pickering and Garrod (2013a). This account is broadly compatible with other integrated accounts, which typically relate to language learning or distribution as well as language processing, such as the *P-Chain* framework (Dell and Chang, 2014) and the *Production-Distribution-Comprehension* account (MacDonald, 2013), as well as by evidence that prediction during comprehension engages production processes (Federmeier, 2007).

To predict the speaker's utterance, we proposed that the addressee attempts to determine the speaker's intention and uses that intention to predict what the speaker would say. For our purposes, the two aspects of this account that we need to consider are (1) that the addressee combines interpretation of the context and covert imitation of the speaker's prior utterance to estimate the intention; and (2) that the addressee uses the intention to predict the speaker's completion in the same way that the addressee would predict his or her own utterance if speaking at that point (though adjusting for differences between the speaker and the addressee). This process is known as *prediction-by-simulation* and works because the comprehender has similar representations and mechanisms to the producer. (Comprehenders may also use *prediction-by-association*, which relies on past experiences during comprehension; see Pickering and Garrod, 2013b, for discussion).

Consider a situation in which a mother is cooking dinner and her son comes into the kitchen and turns to speak. Based on the context (the food, the time, knowledge of her son's habits) but without any utterance, she estimates that his intention is to ask what is for dinner. But he then says *What are we going to do after* … and she now combines the context and the utterance to derive an (updated) intention—that he is producing a question in which the only missing element is something referring to dinner. Pickering and Garrod (2013a) assume that she represents his intention and that this constitutes her own "production command," which sets off the processes that she would use to complete the utterance herself (adjusting for differences between herself and her son). This means that she converts the prior utterance into a production representation via "covert imitation," which is then compatible with the format of the intention.

To understand how addressees predict speakers' utterances, we first note that Pickering and Garrod (2013a) argued that speakers predict their own utterances, using so-called *forward models*. For example, it may take several hundred milliseconds to start naming an object (e.g., Piai et al., 2011), but well before this, speakers can construct representations of what they believe they will say and what they will experience themselves saying. Psycholinguistic evidence for this claim comes from the finding that speakers are affected by the contextual probability of a target word or phrase given the preceding context. If the probability is higher, the speakers are more likely to produce a reduced form (Aylett and Turk, 2004) or to omit an optional word such as the complementizer *that* (Jaeger, 2010). This suggests that the speaker is sensitive to the probability of the target given the context, before uttering the target, and therefore has predicted the target by this point.

Pickering and Garrod (2013a) based their account on the mechanisms of action control, in which people predict movements before they occur and while they are occurring (and use their predictions to make corrections on-line; e.g., Wolpert, 1997). It assumes that people learn the relationship between their intentions and the outcomes (e.g., speech or arm movement), so that the forward model can be computed independently of the implementation of the action. It also assumes that people represent the inverse model of this relationship between outcomes and intentions on the basis of the forward model. They can then use the paired forward-inverse models to predict the outcomes of their actions (via forward models) and subsequently modify those actions when necessary (via inverse models), with both the learning and the on-line control being driven by prediction error minimization. Theories of speech production make such claims about syllables and phonemes (Hickok et al., 2011; Tourville and Guenther, 2011). Pickering and Garrod (2013a) make the more general claim that speakers can concurrently predict at the full range of linguistic levels, such as semantics, syntax, and phonology, and that they also make predictions about timing.

Following this, Pickering and Garrod (2013a) argued that comprehenders predict other people's utterances, again using forward-inverse model pairings. For example, if they believe that their partner is about to name an object, they can construct representations of what they believe their partner will say and what they will experience their partner saying. This is compatible with theories of action perception, in which people predict their partner's unfolding movements (Wolpert et al., 2003; Oztop et al., 2005). To do this, Pickering and Garrod argued that comprehenders covertly imitate the speaker, derive the (putative) intention of the speaker (using a combination of context and inverse model), use that intention to derive their upcoming intention, and treat this upcoming intention as the input to the forward models that predict the upcoming utterance, again at different linguistic levels (see also Pickering and Garrod, 2014). This proposal means that predicting another person's utterance involves the same predictive mechanism used to predict one's own utterance.

Pickering and Garrod (2013a) explained dialog as a form of joint action in which both interlocutors predict both their own and their partner's utterances. The addressee can predict the speaker's unfolding utterance and how he might respond

to that utterance. The speaker similarly can predict how she will continue and how her partner might respond. Well-aligned interlocutors (Pickering and Garrod, 2004) tend to make the same predictions as each other. Moreover, Pickering and Garrod (2014) proposed that interlocutors monitor the quality of these predictions and use the discrepancies between predicted and actual utterances (by themselves and their partners) to control the flow of the dialog.

We propose that interlocutors make two different types of prediction during comprehension, relating to content and timing. The basis for content prediction is the processes of language comprehension typically investigated in psycholinguistics, and involves the extraction of phonology, syntax, and particularly semantics that can be derived from the speaker's utterance. From these representations, the comprehender can predict the phonology, syntax, and semantics of the upcoming utterance. The basis for timing prediction is the speaker's speech rate, which the comprehender can use to predict the rate of the upcoming utterance. We propose that these mechanisms are distinct, but that they can influence each other and be combined for various purposes. We now demonstrate how they can be used to predict turn-endings. At the end of the paper, we illustrate how they can be combined for other purposes, for example to resolve ambiguities (e.g., Dilley and Pitt, 2010).

To return to our example, the mother uses context to determine the boy's putative intention before he starts to speak and predicts that he will produce a fairly short question asking about what is for dinner. After the boy begins to speak, she revises her prediction by combining context with her covert imitation of the boy's incomplete utterance *What are we going to do after* … (a process that is in fact informed by her monitoring the discrepancy between his incoming utterance and her prior prediction). She therefore covertly imitates the boy's utterance, derives the boy's intention in producing *What are we going to do after* and derives her belief about his upcoming intention, which we assume is to produce the word *dinner* and then stop. She then predicts aspects of the form of *dinner* (e.g., main meal, noun, /dɪnər/, rising intonation, two syllables).

Note that Pickering and Garrod (2013a) argued that forward models are likely to be impoverished—not containing all of the information included in the implemented representations underlying actual speech (see several commentaries and Pickering and Garrod, 2013b, for discussion). By repeatedly producing utterances as a result of intentions, the speaker learns different intention-utterance regularities. She can draw on different regularities depending on the situation—for example, predicting the semantic class of the upcoming word (e.g., when predicting whether a speaker is going to suggest one of a set of restaurants) or the initial sound (when predicting whether the speaker is going to suggest a particular restaurant, e.g., *Kalpna*). The speaker predicts different aspects of the upcoming utterance on different occasions. Such flexibility clearly makes the forward models more useful for aiding fluency, but it also means that we cannot determine which aspects of an utterance will be represented on a particular occasion. In Alario and Hamamé's (2013) terms, we assume that the "opt-out" is circumstantial rather than systematic. For example, predictions may contain "fine-grained phonetic detail," contra Trude (2013); see Pickering and Garrod (2013b, p. 379).

Quite separately, she determines his speech rate, which we assume is in terms of syllables, say 170 ms/syllable. Below, we discuss evidence both that speakers compute speech rate in terms of syllables and that they entrain on syllable rate. The boy's mother therefore assumes (without further computation) that the upcoming speech rate will also be 170 ms/syllable. Let us assume that her "target" is to leave a one-syllable gap between her son's contribution and her own (corresponding to what Schegloff, 2000, calls a *beat*). To determine point of initiation, she therefore estimates the length in syllables of her son's predicted completion (2) plus the gap (1), and multiplies them by syllable time (i.e., $3 \times 170\,\text{ms} = 510\,\text{ms}$). At the same time, she constructs linguistic representations for *What are we going to do after dinner* (i.e., including *dinner*), and uses them to prepare an appropriate response (e.g., *Play football*, which is syntactically and semantically appropriate). This preparation involves extension of the forward model to incorporate self- as well as other-prediction, and also involves the implementer—in other words, actual accessing of linguistic representations such as the lexical entries for *play* and *football*. This allows her to utter *Play football* after a one-syllable interval, assuming that he does utter *dinner* and takes 340 ms to do so.

Comprehenders might predict their partner's penultimate word and final word (both in terms of timing and content). Making these two predictions at the same time does not lead to resource competition because they are two compatible predictions, as they follow from the same process of covert imitation: one is the result of production command that would be used to predict the next word [$i_B\ (t+1)$ in the terms of Pickering and Garrod, 2013a], and the other the result of production command that would be used to produce the word after that ($i_B\ (t+2)$). For example, they might predict a completion of *after* (in 340 ms) and *dinner* (in 640 ms). They do not compete for resources. We have also noted that comprehenders make predictions about their partner's completion and their own response (though of course they need to "tag" whether a specific prediction is about themselves or their partner). For example, they might predict their partner's final word *dinner* in 340 ms and their own response *Play football* in 510 ms. If these predictions are compatible, they will also not compete for resources. This will be true if the comprehender is well-aligned with the speaker, something that is likely to be the case in a simple question-answer case such as this. Of course, if someone is trying to comprehend a speaker while preparing an unrelated utterance (e.g., at a "cocktail party"), the self- and other-predictions are unlikely to be aligned and processing difficulties may ensue.

Note also that comprehenders may use forward models to predict multiple alternatives, weighted according to their likelihood (e.g., Wolpert and Kawato, 1998). Such multiple predictions are particularly valuable during comprehension, because the speaker may often produce one of many alternatives (e.g., *dinner, supper, the meal*). In fact, there is some evidence for parallel prediction in both ERP studies (DeLong et al., 2005) and corpus-based investigations of reading time

(Smith and Levy, 2013). Such parallel prediction does not appear to be resource-intensive (as it is in many dual tasks).

Importantly, the content and timing predictions are combined, but they remain separate predictions. The comprehender does not construct a single (indivisible) representation of timing and content. This means that the comprehender can change either timing or content as necessary. For example, the boy might not stop after *dinner* but produce further words, or perhaps speak slowly or disfluently. If so, the mother would need to alter prediction of timing but not content. Alternatively, the boy might (unexpectedly) say *swimming* rather than *dinner*, in which case the mother would have to revise her interpretation (based on monitoring; Pickering and Garrod, 2014) but not timing. Below we explain how the flexibility induced by separate representations appears to be used in practice.

In more general terms, then, we assume that the addressee is constantly covertly imitating the speaker, and uses the process of covert imitation to make predictions about both the timing and the content of the speaker's utterance. This process supports alignment (Pickering and Garrod, 2004), so that the addressee's linguistic representations become more similar to those of the speaker, as well as entrainment of timing (see below). Sometimes the addressee predicts that the speaker is about to finish and that it would be appropriate for the addressee to take the floor. Alongside this, the addressee uses forward modeling to predict the speaker's concluding utterance and the addressee's own response (in a way that is aided by the alignment that has taken place). After the speaker finishes, and assuming that the addressee's prediction is correct or sufficiently close, the addressee speaks appropriately and at the appropriate time. We now discuss how entrainment of timing can take place, before turning to the question of how the addressee monitors the speaker's utterance and how difficulties can be managed.

How does the Addressee Entrain Timing with the Speaker?

Arnal and Giraud (2012) argued that the brain implements predictions about timing and content in different ways. More specifically, predictions about the timing of sensory events are based on cortical oscillations in the low frequency range (*delta* band, 1–3 Hz; *theta* band, 4–8 Hz), whereas predictions about sensory content are based on higher frequency cortical oscillations (*gamma* band, about 30–60 Hz). Both auditory and pre-motor cortex reveal ambient neural oscillations in the *theta* range (Giraud et al., 2007). Those in the auditory cortex become entrained to theta oscillations in the speech envelope (see Gross et al., 2013; Zion Golumbic et al., 2013). These theta oscillations correspond to the frequency of the speaker opening and closing her mouth and hence the rate of her syllabic articulation (Chandrasekaran et al., 2009). According to Arnal and Giraud, predictive timing arises from this low-level mechanism of neural entrainment. In the presence of a fast speaker, the auditory cortex first adapts by increasing the rate of oscillations. These entrained oscillations then become predictive by creating periodical temporal windows for higher-order regions to read out encoded information (see also Kotz and Schwartze, 2010; Giraud and Poeppel, 2012). In other words, low frequency cortical oscillations come to predict the precise timing of critical speech events (at the level of the beginning and end of syllables).

There is now considerable empirical support for this with respect to speech perception. For example, Zion Golumbic et al. (2013) recorded ECoG (Electrocorticographic) activity in the auditory cortex as listeners attended to one of two speakers in a simulated "cocktail party" situation. They found that both the phase of low frequency cortical activity (i.e., delta and theta band) and the power of higher frequency cortical activity (high gamma) tracked the low frequency aspects of the speech envelope (i.e., the speech wave), for the attended but not the unattended speech. Follow-up analyses indicated that the higher frequency effects reflected evoked responses, whereas the low frequency effects reflected processes more closely related to perception. This latter finding suggests that low frequency speech tracking serves to limit the transfer of sensory responses to higher-order brain regions. As the low frequency phase of the attended and unattended speech is likely to be different, the listener can use phase tracking for selective attention.

Furthermore, they found that the precision of low frequency tracking increased steadily from the beginning to end of each attended utterance, consistent with a predictive process. Using a somewhat different approach, Gross et al. (2013) compared oscillatory MEG (Magnoencephalographic) signals in the cortex with those in the speech envelope for a 7-min narrative played both forwards and backwards. Mutual Information analyses revealed that low frequency (i.e., delta, theta band) cortical oscillations (in the right hemisphere auditory cortex) encoded the phase of low frequency oscillations in the speech envelope, whereas higher frequency (i.e., gamma) cortical oscillations (in the left hemisphere auditory cortex) encoded the energy of higher frequency oscillations in the speech envelope. Notably, the degree of oscillatory entrainment was much greater for the forward as opposed to the backward speech. Further, analyses of the forward speech established that transients (i.e., high energy bursts of sound) at the beginning of utterances reset the phase of low frequency cortical oscillations to bring it into line with the phase of low frequency oscillations in the speech envelope. Such resetting of the transients did not occur to the same extent for backward speech. This suggests that these effects reflected top-down predictive processing as opposed to bottom-up evoked responses.

These and related findings (see Ding and Simon, 2014) clearly implicate a low frequency oscillatory tracking system which represents current speech rate and predicts how it will unfold in the immediate future. Although the functional explanation for this entrainment process has been primarily related to syllabic parsing (Ghitza, 2011, 2013) or selective attention (Ding and Simon, 2012; Zion Golumbic et al., 2013), we propose that it may also play an important role in predicting when an interlocutor's turn will end and timing the addressee's response onset.

We hypothesize that cortical theta oscillations entrained during speech comprehension also influence the rate of speech production, probably mediated by mid-brain circuitry (see Giraud et al., 2007; Kotz and Schwartze, 2010). In other words, theta oscillations in auditory cortex entrain theta oscillations in premotor cortex, which in turn influence both the timing of the speech onset and rate of articulation. We assume that the rate and phase of oscillation play a causal role in such entrainment (though it is conceivable that entrainment results from some underlying pattern of neural activity that is highly correlated with oscillation). This is pertinent because (as we have noted), turn-transition involves more than detecting when an interlocutor's turn will end; it also involves initiating one's own turn in a timely fashion, with such inter-turn intervals reflecting the current speech rate. The finding that, during dialog, interlocutors' speech rates and turn transition times become entrained (Street, 1984) is consistent with coupling between the current speaker's rate and the subsequent speech rate of their partner (cf. Jungers and Hupp, 2009, for priming of speech rate in monolog). It is also consistent with Wilson and Wilson's (2005) proposal that the timing of turn-transitions is based on an underlying entrainment of syllabic speech-rate oscillations. Our proposal, therefore, is that interlocutors entrain theta oscillations in auditory cortex and premotor cortex, and that such entrainment underlies the coordination of comprehension and production in turn-taking.

In conclusion, interlocutors entrain their speech rates based on low-frequency acoustic information. This process appears to be quite separate from the mechanisms of prediction-by-simulation and alignment, which are based on linguistic representations. However, the addressee can combine the results of entrainment (i.e., prediction of timing) with those of linguistic prediction (i.e., prediction of content) to determine the appropriate timing for turn transitions, as we illustrated in the previous section (see **Figure 1**).

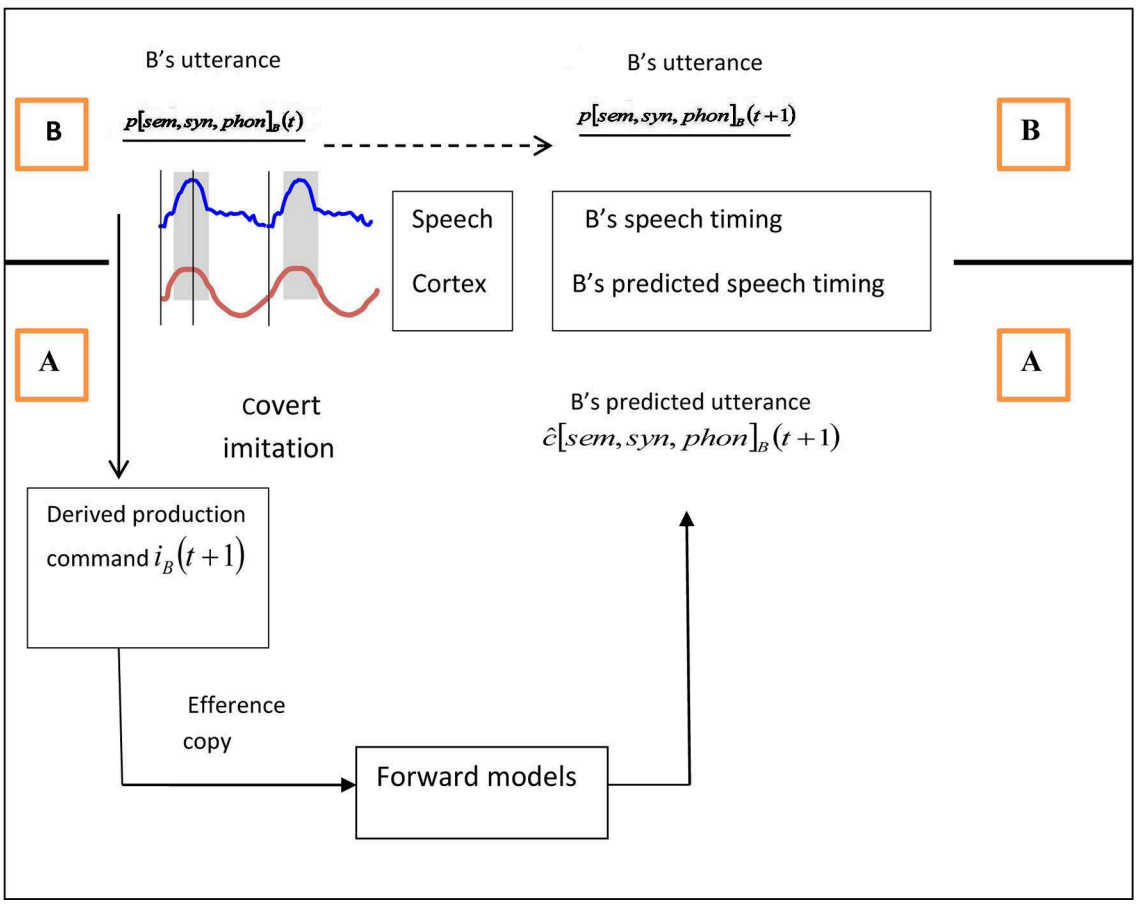


FIGURE 1 | A schematic illustration of the turn ending prediction mechanism, with *A* as addressee (below black line) and *B* as current speaker (above black line). Above the line, *B's* unfolding utterance content is shown as $p\,[sem, syn, phon]_B\,(t)$ and $p\,[sem, syn, phon]_B\,(t+1)$, which refer to semantic, syntactic, and phonological representations of the current utterance (at time *t*) and the upcoming utterance (at time $t+1$, with the underlining indicating that they are *B*'s representations; see Pickering and Garrod, 2013a). The timing of *B's* speech is represented in terms of the entrained theta oscillations in *B's* speech envelope. Below the line, *A's* prediction of the content of *B's* unfolding utterance is shown as $\hat{c}\,[sem, syn, phon]_B\,(t+1)$ and *A's* prediction of *B's* speech timing is shown in terms of theta oscillations in *A's* auditory cortex. The predicted content comes from *A* covertly imitating *B's* utterance at time *t*, deriving *B's* putative production command at time *t*+1 and then feeding this production command into forward models to generate the predictions for time *t*+1. The predicted timing comes from entrainment of *B's* cortical theta oscillations with theta oscillations in *A's* speech envelope.

But because the mechanisms of prediction of timing and content are shared with production, we propose that they also aid the addressee's own utterance production. The form of the question (e.g., *What are we going to do after dinner?*) requires a type of answer (verb phrase specifying an activity), which the addressee can prepare by using the same mechanisms that he uses in comprehension. The addressee's onset and rate of articulation follow from the entrainment of speech rate, and specifically the suggestion that such entrainment may also occur in pre-motor cortex (Giraud et al., 2007). This entrainment could therefore be directly applied to the onset and timing of syllable production in relation to the addressee's response, on the assumption that ambient theta oscillations in pre-motor cortex influence the timing of speech articulation.

Preparing an Appropriate Response

So far we have concentrated on prediction of content and timing of a partner's current contribution and how this enables the addressee to estimate when the turn will end. But addressees do not merely have to predict content and turn endings; they also have to prepare an appropriate response, or decide not to do so. Recent research has begun to consider the extent to which a responder's planning overlaps with the previous utterance. These studies make use of dual-tasking paradigms (e.g., target tracking or finger tapping) to demonstrate more disruption during production than comprehension (Boiteau et al., 2014; Sjerps and Meyer, 2015). Importantly, the main indication of difficulty during comprehension occurs in about the last half second of the previous utterance, suggesting that planning occurs quite late but is time-locked to turn-ending.

To respond appropriately, the new speaker has to determine the speaker's speech act. For example, a non-rhetorical question mandates a (relevant) answer (or some other valid response such as a query), whereas a rhetorical question does not. Because utterance planning takes time (as we have argued), fluid conversation requires that the addressee should (in general) determine the speech act before the utterance is complete. On occasion, it may not be possible to determine the speech act before the end of the utterance (e.g., because the only relevant information is rising intonation, indicating a question). However, such cases are almost certainly quite rare (Levinson, 2012). For example, the widespread occurrence of *Wh*-words or subject-verb inversion (e.g., *Is the* …) as the beginning of a question provides the addressee with a clear early indication of the speech act. In addition, dialog is full of "pre-sequences" (Schegloff, 1988) that make the upcoming speech act clear well in advance (e.g., *Can I ask you a question?*). Of course, responses are generally congruent with the prior utterance. This is obviously the case for semantics, but is also often true for syntax, as in question-answer pairs (e.g., Levelt and Kelter, 1982) or cross-speaker completions (e.g., Clark and Wilkes-Gibbs, 1986). We propose that comprehenders can make use of this congruency when planning their responses, and use it to share resources between comprehension and preparing production, in accord with Pickering and Garrod (2013a) and in particular the prediction-by-simulation route.

How Addressees Take the Floor

We have argued that addressees primarily use prediction-by-simulation to predict the content of the speaker's utterance and use prediction-by-simulation in combination with oscillatory entrainment to predict its timing. Prediction of content is enhanced by alignment at many linguistic levels and also facilitates the formulation of an appropriate response. Prediction of timing is used to determine when the speaker will end, and more importantly, when the addressee should start speaking. We now illustrate our account with examples of speaker-addressee turn transition, some of which include difficulties. As the examples show, turn-transition can be entirely straightforward, but very often it leads to minor disruption that can be internally managed (i.e., by the interlocutors themselves); our focus is on spontaneous conversation. Rather more occasionally, it leads to some form of conversational breakdown.

First consider an excerpt from (1) above. Bee describes purchasing an expensive art book and then produces *you know* (highlighted). While hearing this, Ava predicts that Bee is likely to end at this point and that Ava can (or should) take the floor (i.e., this constitutes a potential turn-transition point). The timing of the response is the result of entrainment based on Bee's speech rate. Ava's response *not drop it* reuses part of Bee's previous utterance, as expressed in the words *drop* and *it* and the way they are combined; this repetition occurs because Ava has linguistically aligned with Bee. Interestingly, Bee speaks at the same time as Ava, and produces a semantically equivalent utterance (*hold onto it*). This shows that both Ava's utterance and her timing were appropriate and that her prediction was successful.

1 (excerpt).
Bee: Yeuh he-ez he wz handing me the book en 'etol' me twunny dolliz I almos' dro(h)pped i(h)[t 'hh 'hh
Ava: [thhunh.
Bee: 'hhh I said but fer twunny dollars I bettuh hh 'hh **yihknow**,
(0.2)
Bee: 'hhh h[**hold o:nto it** (h)hh] huhh huh] 'hh!
Ava: [**not drop it**.] huhh huh]
P(0.2)
Bee: Ih wz, (0.2) y'know (fun)....

However, Bee's response also creates a problem, because it means that Bee wishes to continue speaking. Ava and Bee's overlap is quite extensive, presumably because they are semantically well aligned (and may therefore find it possible to comprehend and produce three-word overlaps). But they then both produce laughter and stop speaking, before Bee continues. In terms of Pickering and Garrod (2014), after *you know*, Ava predicts that she will say *not drop it* after (say) 300 ms (corresponding to the silence plus laughter). This self-prediction turns out to be correct. Although she may realize what Bee would have said at this point, she presumably does not predict that Bee will also speak at this time, as overlapping speech is strongly disfavored. When Bee does speak, Ava compares her prediction that Bee will not speak with the actual event. This leads to a conflict that could result

in her stopping speaking, but in fact she judges that uttering the three words will not be problematic. Similarly, Bee presumably does not predict that Ava will speak at that point, but also judges that continuation would not be problematic. However, the overlap between Ava and Bee can be seen analogous to a speech error (i.e., internal to one speaker) and the laughter, pause, and Bee's eventual continuation can be seen as a form of repair [see points (a) and (g) in discussion of Sacks et al., 1974].

In other cases, the transitions are not quite so successful and require some management. In Example 2 (from Schegloff, 1996, p. 85), Ava begins by describing her unexpected activity[3].

2.

Ava: I'm so:: ti:yid. I j's played ba:ske'ball t'day since the firs' time since I wz a freshm'n in hi:ghsch[ool.]
Bee: [Ba:]sk(h)= =etb(h)a(h)ll? (h)[(•owhe(h)re
Ava: [Yeah fuh like an hour enna ha:[lf.]
Bee: [.hh]
Bee: Where didju play ba:sk[etbaw.]
Ava: [(The) gy:m].
Bee: In the gy:m?

Bee appears to predict that Ava is in the middle of uttering *high school* and about to finish speaking (or at least, reach a turn-transition point). Bee therefore queries *Basketball?*, indicating surprise. Ava appears to interpret Bee's contribution as providing an invitation for Ava to expand, but in fact Bee intends to continue with a more specific question (beginning *where*). So both Ava and Bee predict their own utterance and its timing, but also predict that their partner is not about to speak. When their partner does speak, a clash ensues—with Ava continuing but Bee ceding the floor. However, we propose that Bee retained her question (i.e., her planned utterance) until she was able to predict a turn-transition point (toward the end of *half*) and then produced it. Ava, in turn, predicted that Bee was uttering *basketball* after the first syllable and produced an appropriate response (*the gym*) as Ava was finishing her question. We propose that dealing with these transitions requires speakers to make separate predictions of both content and timing, in accord with our account. (Note that the overlap is twice associated with minor disruption to the first speaker's turn ending, both in *high school* and *basketball*; see Schegloff, 2000).

So far, our examples have been from dyadic interactions. In multi-party conversations, different addressees may be permitted to speak at a turn-transition point. This situation can often lead to short periods of overlapping speech (where there is "competition for the floor"). In (3), Kathy is describing hand-weaving at a dinner party (from Schegloff, 2000, p. 31). After pausing and saying *you know*, both Dave and Rubin speak at the same time. Dave withdraws, and Rubin completes a question to Kathy, who responds to him. We propose that both Dave and Rubin predict that Kathy is about to complete her utterance *you know* and that a response is appropriate. They both predict timing correctly, so that they start immediately after Kathy finishes and therefore at the same time as each other. But neither predicts that the other is about to speak. Hence, there is a large discrepancy between their predictions and what actually happens. (Of course, it is possible that Dave predicts that Rubin would speak but decided to speak anyway, in which case Dave would not encounter such a large discrepancy during monitoring). Dave's approach to this discrepancy is to abandon speech, thus preventing the communicative failure that would likely occur following extended overlap, whereas Rubin's approach is to carry on regardless (perhaps assuming that Dave will give up)[4].

3.

Kathy: So once I'd set up the warp, i' w's very simple to jus' keep-jus' to weave it.
(0.8)
Kathy: You know[( )
Dave: [(But listen tuh how long)]
Rubin: [In other words, you gotta string up thee] you gotta string up thee colors, is that it
Kathy: [Right]
Rubin: [in thee] in thee [warp.]
Kathy: [right]

Finally, we note that the addressee can separate the process of prediction from the process of preparing a response. The response can be "ready" before it is executed (just as in Ferreira, 1991; Piai et al., 2011). In (4) (from Schegloff, 2000, p. 25), a family is querying Anne's claim that she used to buy a pair of shoes a month before she was married, and her husband Dick keeps attempting to make a joke about it:[5]

[3] A more literal rendition of this exchange is:
Ava: I'm so tired. I just played basketball today. The first time since I was a freshman in high school –
Bee: Basketball – where?
Ava: Yes, for like an hour and a half-
Bee: Where did you play basketball?
Ava: The gym
Bee: In the gym?

[4] A more literal rendition of this exchange is:
Kathy: So once I have set up the warp, it was very simple to just keep – just weave it. You know.
Dave: But listen to how long –
Rubin: In other words, you have got to string up the – you have got to string up three colors, is that it.
Kathy: Right.
Rubin: In the – in the warp.
Kathy: Right

[5] A more literal rendition of this exchange is:
Anne: Every six months I went in for shoes. and I had– must have had about, a hundred pairs a shoes.
Deb: Really mother you spent–
Dick: You know what -
Deb: Boy were you wasted
Dick: you know she exaggerated slightly.
Dick: You know what– you know–
Deb: What a waster you were
Anne: Don't say that I'm exa– just say I'm a liar.
Dick: You know what your –
Deb: It's not a question of lying it's a question of being–
Dick: Your grandmother is a centipede that's why she has to have a hundred pairs of shoes.

4.

Anne: Every six months I wen' in fih shoes. 'n I had-must'v had about, (0.5) a hundred pairs a shoes.
(2.0)
Deb: Really mother = you spent-
(1.0)
Dick: **You know [wha : t,]**
Deb: [Boy we]re you:: w- [w a s t e d]
Dick: [(you know) sh-]
Exaggerated slightly.
(0.8)
Dick: **Y' [know what-y'know- [( )]**
Deb: [w h a t a w a s [ter you] w e r e]
Anne: [DON'T S]AYthat I'm ex]=
=a- just say I'm a liar.
Dick: **Y'know what, yer** [grandmother -] =
Deb: [>'ts nota question<] of =
=[<ly:ing 't's a question of being- >]
Dick: =[yer GRANDMOTHER IS A CENTI]PE:DE,
that's why- she esstuh hev a khundred pairs of shoes.

At (24), (29), and (33) (in bold), Dick tries and fails to utter the joke (i.e., Anne is a centipede) that he eventually manages at (36). Dick uses prediction of timing and content to determine a turn-transition point on all four occasions. However, Deb manages to capture the floor three times. On each occasion, Dick has a prepared utterance, which is presumably ready throughout the interchange (from [24] onwards, at least), and hence the preparation of the utterance is separate from the predicted timing. This is a further indication of the distinction between mechanisms for timing and content.

Implications and Discussion

We have shown how content can be combined with timing to predict the end of the interlocutor's turn and determine the appropriate moment to speak. But content and timing can also be used to determine content itself. A good example comes from Dilley and Pitt (2010), who presented listeners with a context spoken at different rates preceding the phrase *leisure or time* and found that they tended to hear it as *leisure time* (i.e., without *or*) if the context was spoken slowly. They then presented listeners with a context preceding the phrase *leisure time* and found that they tended to hear the phrase as *leisure or time* if the context was spoken quickly. Presumably, participants are entrained to the contextual speech rate and then predict that the upcoming phrase will also be produced at that rate. Their interpretation of the phrase is therefore dependent on their predictions. In terms of **Figure 1**, the predicted timing is used to help determine utterance content.

In this paper, we have focused on the role of prediction during comprehension on turn transition. Specifically, we have argued that comprehenders predict the speaker's content and speech rate, and use these to compute what they are likely to say and how quickly they are likely to say it. We also assume that such prediction helps the comprehender decide when to speak and what to say. However, Pickering and Garrod (2013a) also proposed that prediction during comprehension aids comprehension itself (e.g., facilitating word recognition in noise), aids learning (as comprehenders learn from the discrepancy between the prediction and the actual speech), permits other monitoring (e.g., detecting speaker's errors; Pickering and Garrod, 2014), and assists in the process of alignment (Pickering and Garrod, 2004). Finally, we note that our account is consistent with the effects of timing disruption in dialog. It has been known for 50 years that delaying transmission can seriously disrupt conversation (e.g., Krauss and Bricker, 1967).

A specific set of empirical predictions following from this account concern the separation of timing and content. In a turn-taking paradigm (e.g., question-answering), there should be separate effects of content difficulty (e.g., hard vs. easy questions) and regularity of timing (e.g., varying regularity of speech rate). But in addition, we propose that turn-taking relates to a combination of timing and predicted length in syllables. If a speaker expects a long sentence-final word but gets a short one (e.g., *Is the largest animal in zoo the bear?*, when *elephant* is expected), then the turn interval should be larger than if the expected word was short (*Is the fiercest animal in the zoo the bear*, when *lion* is expected), but this interval should also be affected by speech rate. Experiments such as these should be able to show how predictions of timing and content are separable but ultimately combined in turn-taking.

In conclusion, we have presented a cognitive account to explain the skill with which conversationalists manage turn-transitions in dialog. The account covers addressees' ability to predict when their interlocutor's turn will end, to craft an appropriate response, and to implement the response in a timely fashion. To do this, we propose that they make use of prediction-by-simulation to predict upcoming content and oscillatory entrainment to predict timing. Whereas predicted content depends on forward modeling mechanisms similar to those used in control of speech production, predicted timing results from sensitivity to characteristics of the speech envelope. However, the addressee brings these predictions together in a way that leads to well-coordinated dialog, with very brief turn transitions. In this way, we propose that interlocutors are able to make an apparently difficult aspect of conversation appear remarkably straightforward.

Acknowledgments

Martin J. Pickering acknowledges support from a Leverhulme Trust Research Project Grant, RPG-2014-253.

References

Alario, F.-X., and Hamamé, C. M. (2013). Evidence from, and predictions from, forward modeling in language production. *Behav. Brain Sci.* 36, 348–349. doi: 10.1017/S0140525X1200249X

Arnal, L. H., and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398. doi: 10.1016/j.tics.2012.05.003

Aylett, M., and Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang. Speech.* 47, 31–56. doi: 10.1177/00238309040470010201

Boiteau, T. W., Malone, P. S., Peters, S. A., and Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *J. Exp. Psychol. Gen.* 143, 295–311. doi: 10.1037/a0031858

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5:e1000436. doi: 10.1371/journal.pcbi.1000436

Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* 22, 1–39. doi: 10.1016/0010-0277(86)90010-7

Dell, G. S., and Chang, F. (2014). The P-Chain: relating sentence production and its disorders to comprehension and acquisition. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369:20120394. doi: 10.1098/rstb.2012.0394

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504

De Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

Dilley, L. C., and Pitt, M. A. (2010). Altering context speech rate can cause words to appear and disappear. *Psychol. Sci.* 21, 1664–1167. doi: 10.1177/0956797610384743

Ding, N., and Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89. doi: 10.1152/jn.00297.2011

Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *J. Mem. Lang.* 30, 210–233. doi: 10.1016/0749-596X(91)90004-4

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2:130. doi: 10.3389/fpsyg.2011.00130

Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* 4:138. doi: 10.3389/fpsyg.2013.00138

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frachowiak, R. S. J., and Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56, 1127–1134. doi: 10.1016/j.neuron.2007.09.038

Giraud, A. L., and Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat. Neurosci.,* 15, 511–517. doi: 10.1038/nn.3063

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzerri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11:e1001752. doi: 10.1371/journal.pbio.1001752

Hale, J. (2001). "A probabilistic early parser as a psycholinguistic model," in *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (Pittsburgh, PA).

Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration of speech processing: computational basis and neural organization. *Neuron* 69, 407–422. doi: 10.1016/j.neuron.2011.01.019

Indefrey, P., and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001

Jaeger, F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cogn. Psychol.* 61, 23–62. doi: 10.1016/j.cogpsych.2010.02.002

Jungers, M. K., and Hupp, J. M. (2009). Speech priming: evidence for rate persistence in unscripted speech. *Lang. Cogn. Process.* 24, 611–624. doi: 10.1080/01690960802602241

Kotz, S. A., and Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortical-cortical framework. *Trends Cogn. Sci.* 14, 392–399. doi: 10.1016/j.tics.2010.06.005

Krauss, R. M., and Bricker, P. D. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. *J. Acoust. Soc. Am.* 41, 286–292. doi: 10.1121/1.1910338

Levelt, W. J. M., and Kelter, S. (1982). Surface form and memory in question answering. *Cogn. Psychol.* 14, 78–106. doi: 10.1016/0010-0285(82)90005-6

Levinson, S. C. (2012). "Action formation and ascription," in *Handbook of Conversational Analysis,* eds J. Sidnell and T. Stivers (Oxford: Blackwell), 103–130.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177. doi: 10.1016/j.cognition.2007.05.006

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Front. Psychol.* 4:226. doi: 10.3389/fpsyg.2013.00226

Magyari, L., and De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

Oztop, E., Wolpert, D., and Kawato, M. (2005). Mental state inference using visual control parameters. *Cogn. Brain Res.* 22, 129–151. doi: 10.1016/j.cogbrainres.2004.08.004

Piai, V., Roelofs, A., and Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: attention and task decisions. *J. Mem. Lang.* 64, 404–423. doi: 10.1016/j.jml.2011.01.004

Pickering, M. J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–225. doi: 10.1017/S0140525X04000056

Pickering, M. J., and Garrod, S. (2013a). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–392. doi: 10.1017/S0140525X12001495

Pickering, M. J., and Garrod, S. (2013b). Forward models and their implications for production, comprehension and dialogue. *Behav. Brain Sci.* 36, 377–392. doi: 10.1017/S0140525X12003238

Pickering, M. J., and Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Front. Hum. Neurosci.* 8:132. doi: 10.3389/fnhum.2014.00132

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., and Halgren, E. (2009). Sequential processing of lexical, grammatical, and articulatory information within Broca's area. *Science* 326, 445–449. doi: 10.1126/science.1174481

Schegloff, E. A. (1988). Presequences and indirection: applying speech act theory to ordinary conversation. *J. Pragmat.* 12, 55–62. doi: 10.1016/0378-2166(88)90019-7

Schegloff, E. A. (1996). "Turn organization: one intersection of grammar and interaction," in *Interaction and Grammar,* eds E. Ochs, E. A. Schegloff and S. A. Thompson (Cambridge: Cambridge University Press), 52–133.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking in conversation. *Lang. Soc.* 29, 1–63. doi: 10.1017/S0047404500001019

Sjerps, M. J., and Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition* 136, 304–324. doi: 10.1016/j.cognition.2014.10.008

Smith, N. J., and Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition* 128, 302–319. doi: 10.1016/j.cognition.2013.02.013

Staub, A., and Clifton, C. Jr. (2006). Syntactic prediction in language comprehension: Evidence from either…or. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 425–436. doi: 10.1037/0278-7393.32.2.425

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universality and cultural specificity in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Hum. Commun. Res.* 11, 139–169. doi: 10.1111/j.1468-2958.1984.tb00043.x

Tourville, J. A., and Guenther, F. K. (2011). The DIVA model: A neural theory of speech acquisition and production. *Lang. Cogn. Process.* 26, 952–981. doi: 10.1080/01690960903498424

Trude, A. M. (2013). When to simulate and when to associate? Accounting for inter-talker variability in the speech signal. *Behav. Brain Sci.* 36, 375–376. doi: 10.1017/S0140525X12002701

Van Berkum, J. J. A., Brown, M. C., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 443–467. doi: 10.1037/0278-7393.31.3.443

Wilson, M., and Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* 12, 957–968. doi: 10.3758/BF03206432

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends Cogn. Sci.* 1, 209–216. doi: 10.1016/S1364-6613(97)01070-X

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. Lond. B* 358, 593–602. doi: 10.1098/rstb.2002.1238

Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw.* 11, 1317–1329. doi: 10.1016/S0893-6080(98)00066-5

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 September 2015
doi: 10.3389/fpsyg.2015.01416

Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model

Laurence Henry[1], Adrian J. F. K. Craig[2], Alban Lemasson[1,3] and Martine Hausberger[1]*

[1] *Laboratoire d'éthologie animale et humaine, Centre National de la Recherche Scientifique, UMR 6552, Université de Rennes 1, Rennes, France,* [2] *Department of Zoology and Entomology, Rhodes University, Grahamstown, South Africa,* [3] *Laboratoire d'éthologie animale et humaine, Centre National de la Recherche Scientifique, UMR 6552, Station Biologique, Université de Rennes 1, Paimpont, France*

Turn-taking in conversation appears to be a common feature in various human cultures and this universality raises questions about its biological basis and evolutionary trajectory. Functional convergence is a widespread phenomenon in evolution, revealing sometimes striking functional similarities between very distant species even though the mechanisms involved may be different. Studies on mammals (including non-human primates) and bird species with different levels of social coordination reveal that temporal and structural regularities in vocal interactions may depend on the species' social structure. Here we test the hypothesis that turn-taking and associated rules of conversations may be an adaptive response to the requirements of social life, by testing the applicability of turn-taking rules to an animal model, the European starling. Birdsong has for many decades been considered as one of the best models of human language and starling songs have been well described in terms of vocal production and perception. Starlings do have vocal interactions where alternating patterns predominate. Observational and experimental data on vocal interactions reveal that (1) there are indeed clear temporal and structural regularities, (2) the temporal and structural patterning is influenced by the immediate social context, the general social situation, the individual history, and the internal state of the emitter. Comparison of phylogenetically close species of Sturnids reveals that the alternating pattern of vocal interactions varies greatly according to the species' social structure, suggesting that interactional regularities may have evolved together with social systems. These findings lead to solid bases of discussion on the evolution of communication rules in relation to social evolution. They will be discussed also in terms of processes, at the light of recent neurobiological findings.

Keywords: turn-taking, vocal interactions, conversation rules, mammals, birdsong, sturnids

OPEN ACCESS

Edited by:
Marisa Casillas, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Daniel Yasumasa Takahashi, Princeton University, USA
Federico Rossano, Max Planck Institute for Evolutionary Anthropology, Germany

****Correspondence:***
Laurence Henry, Laboratoire d'éthologie animale et humaine, Centre National de la Recherche Scientifique, UMR 6552, 263 Avenue du Général Leclerc, 35042 Rennes Cedex, France
lhenry@univ-rennes1.fr

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *01 February 2015*
Accepted: *04 September 2015*
Published: *28 September 2015*

Citation:
Henry L, Craig AJFK, Lemasson A and Hausberger M (2015) Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model. Front. Psychol. 6:1416. doi: 10.3389/fpsyg.2015.01416

Introduction

The Human "bases"

Vocal communication is widespread in the animal kingdom and vocal interactions are an important part of social functioning. Temporal and structural regularities depend on the species' social structure, or may even depend on the immediate context. Two extremes are generally encountered, with either an overlap superposition of acoustic signals between interlocutors or a strict alternation

of vocal utterances: a first emitter leaves a silent interval before producing the following sound during which the second emitter can respond. As in humans, animal vocal interactions may be dyadic ("face to face") or at the group level.

The question though is to what extent these regularities may be functionally convergent with human communication rules, such as turn-taking.

According to Logue and Stivers (2012), the analysis of conversation in humans is based on methods and theories that emerged from sociology in the 70s. One "founder" paper was that of Sacks et al. (1974) who formalized the basics of turn-taking rules and defined them according to eight characteristics (see also Craig and Washington (1986): (1) only one person speaks at a time, (2) the number of participants may vary, (3) the order for speaker turns is variable, (4) turn size is not fixed, (5) the content of speaker turns is spontaneous, (6) simultaneous speech is infrequent and brief, (7) techniques exist for repairing turn exchange errors, (8) turn allocation techniques are used to regulate the exchange. The two major elements are the alternation of utterances between interlocutors and the avoidance of overlap, hence temporal features. In most human cultures, overlap appears as a conversation failure (Sacks et al., 1974) and can lead to the end of the exchange. In human conversation, the fundamental frequency declines, changes in gazing and other subtle signs are used to guide conversational turn-taking (Gérard, 1987; Hauser, 1992). Another important point is that conversational turn-taking rules are acquired during development through adult modeling (Locke, 1993). It is even considered as a child's major achievement, which is made possible by the early stages of parent-child interaction (e.g., Rutter and Durkin, 1987). The mother is seen by some authors as controlling the child rather than facilitating it in the mother-child dyadic interactions (Miura, 1993). Adults may play a major role in canalizing the flow of speech so that it is fragmented enough to allow turns between speakers. Neglected children fail to develop this ability, showing irrelevant turns, interruptions, simultaneous talking and non-contingent responding (Black and Logan, 1995).

According to Calame-Griaule (1965), "In the Dogon society, overlap with someone's speech is a serious impoliteness: these words that could not follow their natural way will be repressed in the spleen. The spleen is the seat of grudge and humiliations. Thus, accumulating repressed words can make sick." Speech has to submit to rules to become an instrument of social communication. In general, "repair mechanisms exist for dealing with errors and violations: stop prematurely or display even rituals." Overall, turn-taking allows interlocutors to enhance mutual attention and responsiveness (France et al., 2001) which may explain why overlapping/interruption is perceived negatively, preventing the other's turn to occur but also indicating a lack of attention.

Overlapping may also reveal a person's status for example. In his work on Kirundi language in Burundi, Albert (1964) found that the order in which individuals speak in a group is strictly determined by seniority of rank: "the rule for servants, females and other inferiors is to speak when spoken to but otherwise to maintain silence in public." Leaders talk more than other individuals (France et al., 2001). Men are more likely to interrupt than women which is generally interpreted as a male "power demonstration" but could also be interpreted as reflecting distinct male and female "subcultures" (Maltz and Borker, 1982). Turn-taking shows a level of contextual adaptation: there is for example variability in turn order, turn size, length of pauses according to the number of individuals present (Sacks et al., 1974).

Apart from simple "politeness," it is obviously difficult to maintain mutual comprehensibility when participants talk at the same time (Duncan, 1972).

In fact, turn-taking is a very general feature of social interactions (games, traffic at intersections …) (Sacks et al., 1974). Turn-taking, "as an orderly distribution of opportunities to participate in social interaction" has been considered, like other such types of interactions, one of the "most fundamental preconditions" for a viable social organization (Schegloff, 2000). According to Sidnell (2001), such rules would correspond to a species-specific adaptation to the contingencies of human interactions, a view shared by Albert (1964) who suggested that this type of interaction is not open to a great deal of cultural diversification. Indeed it is found in a variety of cultures: Thaï (Moerman, 1977), Creols of New Guinea (Sankoff, 1980), Dogon (Calame-Griaule, 1965) amongst others. Although cultural differences are apparent in the duration of pause between turns (the minimal pause under which locutors feel they have been interrupted is about of 0.3 s in France, 0.5 in USA, 1 s in Alaskan Althabascans, Kerbrat-Orecchioni, 2001), avoidance of overlapping appears in all types of languages as well as a minimum gap between turns whatever the languages structure (Stivers et al., 2009). More, within languages, variations in the delay of response are predicted by the same factors such as confirmation or disconfirmation responses or questioner gazing at responder. Indeed, turn taking can be considered as a universal feature within human languages (Stivers et al., 2009).

In all cases, the respect for turn-taking rules requires attention and control and may have evolved over time on the basis of the first rulers who may have been the first to control their vocal production and listen while being listened to MacWhinney (2008). A Dogon saying is that "rules in language = law and order in the society" (Calame-Griaule, 1965).

The universality of turn-taking in humans raises questions about its biological bases and evolutionary trajectory. If it is a species-specific adaptation to social requirements as proposed by Sidnell (2001), there may be either some phylogenetic roots to be found in our closest relatives (non-human primates) or convergence in species with similar social contingencies (Hausberger et al., 2008). Functional convergence (one process of homoplasy) is a widespread phenomenon in evolution, sometimes revealing striking functional similarities between distantly-related species even though the mechanisms involved may be different (Deleporte, 2002). One well known example is birdsong, considered for many years now as the best animal model of language development (Marler, 1970). Amongst the parallels is the observation that both human language and birdsong need to be learned from adult models during development whether in terms of production, perception or usage.

In what follows we will review the turn-taking literature in primates, and the next closest groups which are the non-primate mammals. Then, we will review how birdsong, a flexible type of vocalization, can give rise to vocal interactions whether between group members or territorial or social neighbors.

Coordination in Mammals and Birds' Vocal Interactions

While some aspects of these conversation rules may really be human-specific (e.g., lexical aspects) and difficult to evaluate in animals, other characteristics such as the influence of the number of interlocutors, their social status, "cultural" (interpopulational) differences, the individual's life experience, and the influence of its internal state on the temporal and structural organization of potential turn-taking bouts, can be investigated in animal vocal interactions. Both alternation and overlap (chorus, duets) occur in animal vocal interactions.

In a variety of mammal social species, vocal exchanges occur between a limited number of interlocutors, mostly 2 or 3 (dolphins *Tursiops truncatus*, Janik, 2000; Tyack, 2000; elephants *Loxodonta africana*; Soltis et al., 2005; Campbell's monkeys *Cercopithecus campbelli*, Lemasson et al., 2010).

Non-human primates and other mammals may display the three "classical" forms of temporal organization of vocal interactions (duets, choruses and alternations). Thus, an "organized" *overlapping* is observable when two sperm whales (*Physeter microcephalus*) adjust their timing of "codas" (series of clicks) production (Schulz et al., 2008) or in gibbons who duet by synchronizing their vocalizations (male-female, Geissmann, 2002; mother-daughter, Koda et al., 2013). The extreme case is a chorus where a group joins in calling (e.g., bat spp., Kunz, 1982; Barbary macaques *Macaca sylvanus*, Hammerschmidt et al., 1994; bottlenose dolphins, Kremers et al., 2014; humpback whales *Megaptera novaeangliae*, Au et al., 2000; chimpanzees *Pan troglodytes* Fedurek et al., 2013). *Alternation (antiphony)* is however particularly common in the social call exchanges of different species (bottlenose dolphins, Janik, 2000; elephants, Soltis et al., 2005; Campbell's monkeys, Lemasson et al., 2010; squirrel monkeys, Masataka and Biben, 1987; Diana monkeys, Candiotti et al., 2012; Japanese macaques, Lemasson et al., 2013; bonobos, Touitou et al., in revision; white-winged vampire bats *Diaemus youngi*, Carter et al., 2008; naked mole-rats, Yosida et al., 2007). The structure of sounds is then adapted in that they are often short and produced in sequences with a silent interval, longer than the call itself thus enabling response without overlap. Interval between calls varies according to species (generally 1 s or less but up to 30 s in elephants) and temporal regularities may change within a species: according to call types and their functions (Yamaguchi et al., 2009), to the partner's identity (Biben et al., 1986) and distance (Sugiura, 2007) suggesting an adaptation to the longer latency of response from a more distant partner. The status of the emitter as well as its age are also important for the selectivity of interlocutors within groups. In some species, affiliated individuals exchange more calls (squirrel monkeys *Saimiri sciureus*, Masataka and Biben, 1987; elephants, Soltis et al., 2005; bonobos *Pan paniscus*, Touitou et al., in revision). In other species, the calls of older (Campbell monkeys, Lemasson et al., 2010, Japanese macaques *Macaca fuscata*, Lemasson et al., 2013; marmosets *Callithrix jacchus*, Chen et al., 2009) or higher-ranked (naked mole-rats *Heterocephalus glaber*, Yosida and Okanoya, 2009) individuals will elicit more vocal responses. Individuals can detect and wait for silent windows to vocalize (e.g., cotton top tamarins *Saguinus oedipus*, Versace et al., 2008). This alternation analytic perspective can be extended to non-vocal communication. Gestural signaling sequences can also be considered as interactional projects that develop through courses of action with comparable (<1 s) short delay between requests and responsive moves in both human and non-human primates (Rossano, 2013: Rossano and Liebal, 2014). It has then been proposed that "conversations," following turn-taking rules, could even be detected in non-human primates (Snowdon and Cleveland, 1984; Symmes and Biben, 1988; Hauser, 1992; Lemasson et al., 2010). Thus, pygmy marmosets *(Cebuella pygmaea)* call in sequence more frequently than expected by chance, while the likelihood of an animal calling twice before the other animal called once was less than expected by chance (Snowdon and Cleveland, 1984). These findings clearly demonstrated that the conversation rules were based on social conventions and that the alternation of calling appeared to be adaptive. This was confirmed recently using a coupled oscillator model revealing dynamics such as those proposed for human conversational turn-taking (Takahashi et al., 2013a). In Japanese monkeys and vervets *(Chlorocebus pygerythrus)*, Hauser (1992) described a decrease of the fundamental frequency before ending a call that could "guide" the turns. He estimated that 1/38 calls were interrupted when the exchange was between adult emitters compared to 6/20 were when the individuals were young. This observation suggests that the ability to respect turns may be acquired during development. This was confirmed by Lemasson et al. (2010, 2011) who showed that young primates are 12 times more likely to interrupt turn-taking by calling twice successively than are adults and by Chow et al. (2015) who demonstrated that common marmoset parents guide vocal turn taking development in their young. In humans, self-monitoring is an essential ability for turn-taking, fully developed only after 2 years of age (MacDonald et al., 2012). In a study on parent-infant vocal interactions in marmosets, it was found that only adults have the capacity to self-monitor their vocal output and avoid call overlap (Takahashi et al., 2013b). According to these authors the neural mechanism underlying the development of self-monitoring could be based on the interactions between three neural structures (representing limbic, motor and auditory regions) with feedback connectivity.

In many species, birdsong occurs mostly in Spring at breeding time and is related to territorial defense and mate attraction (Catchpole and Slater, 1995) and conveys information on individual identity, distance, residency (Falls and Brooks, 1975). It also occurs in the winter flocks and at night roosts for the same species, at a time when they gather in larger groups. In social species, song often occurs all year round and is produced in the context of both intragroup and intergroup encounters (e.g., Brown and Farabaugh, 1997).

In territorial songbirds, networks of neighbors, sharing song structures, can be observed; they also react less aggressively to each other than toward a stranger (Falls, 1982; see Catchpole and Slater, 1995; Briefer et al., 2008) forming a "pseudosocial structure." Birdsong has long been considered as a male behavior, but in many cases females do sing more than was thought (Riebel, 2003).

Birdsong interactions present a whole range of temporal modalities: alternation is by far the most common form, but duetting and choruses also occur.

Duetting is considered as a feature of a pair while chorusing is a group activity (Catchpole and Slater, 1995): in white browed sparrow weavers (*Plocepasser mahali*), the dominant male sings a solo, the dominant pair duets, and the group performs **choruses** (Voigt et al., 2006). Duetting can be antiphonal, or overlapping and synchronized (Hooker and Hooker, 1969; Todt and Hultsch, 1982; Trainer et al., 2002). It seems that most duetting species are monogamous, monomorphic, sedentary and that in about one third of the cases, duetting is antiphonal, one third totally overlapping and one third variable between both (Dahlin and Benedict, 2013). In Australian magpies (*Cracticus tibicen*), choruses occur where the whole social group sings together without clear coordination, in particular in the context of intergroup encounters (Brown and Farabaugh, 1991, 1997). Communal singing is one major characteristic of roosting behavior, where choruses occur before the sleeping phase (Counsilman, 1974). The functions of such communal singing have been suggested to be a synchronization of activities, social bonding, and group or territorial defense (Brown and Farabaugh, 1991; Foote et al., 2008).

Alternation is predominant and is based on a singing style that ensures a silent interval after each emission, leaving space for a response (Naguib and Mennill, 2010). In the winter wren (*Troglodytes hiemalis*), 90% of the songs are produced during interactions and the intersong interval is longer when there is a vocal interaction than when the male sings solo (Camacho-Schlenker et al., 2011). Receivers avoid actively overlapping (Wasserman, 1977): in lesser skylarks (*Alauda gulgula*), if two birds start singing simultaneously, one of them stops within 2 s (Gochfeld, 1978, see also nightingales *Luscinia megarhynchos*, Naguib, 1990). In playback experiments, birds often start singing just after the playback in order seemingly to avoid overlapping the next song (Searcy and Beecher, 2011).

Overlapping (one bird starts singing before the other has finished, Todt and Naguib, 2000), may occur during these interactions. In general, it stops the exchange: the first emitter falls silent (Schmidt et al., 2006; Naguib and Mennill, 2010). In black capped chickadees (*Poecile atricapillus*), dominant males tend to overlap more which could reflect increased aggressiveness (Ficken et al., 1978; Baker et al., 2012). In robins (*Erithacus rubecula*) and black capped chickadees, overlapping excites the overlapped interlocutor (Dabelsteen et al., 1997; Mennill and Ratcliffe, 2004). It has been suggested that overlapping may be perceived as a directed aggressive signal (Naguib and Kipper, 2005) or even a signal *per se* (Naguib and Mennill, 2010), but more experimental evidence is still needed (Searcy and Beecher, 2009). Alternation in birdsong exchanges suggests turn-taking rules in that the timing allows turns to be taken between two or more interlocutors, and overlapping elicits "irritation" or a rupture of the exchange. However, we do not know how these characteristics are acquired, what their real significance is and how they are influenced by status or bonding. Social structure may be a key factor.

Alternation requires discontinuous songs that leave space for responses and indeed some "true" territorial species with long continuous songs such as skylarks cannot show this alternating pattern (Geberzahn and Aubin, 2014). Alternation appears more in species with distant vocal interactions but social or "pseudo social" types of relationships. Family or very cohesive social groups are more likely to perform choruses.

Many species, such as caciques *Cacicus* sp. (Feekes, 1982; Thieltges et al., 2014), nightingales (Sorjonen, 1983; Naguib et al., 2002), five striped sparrows *Amphispiza quinquestriata* (Groschupf, 1985), great reed *warblers Acrocephalus arundinaceus* (Catchpole, 1983) have two categories of songs that allow both temporal singing styles: a long continuous (often quiet) vocalization often associated with intersexual interactions, and louder, shorter and simpler songs that are more involved in male-male encounters at a distance (Catchpole and Slater, 1995).

In summary, vocal interactions in animals are clearly regulated, especially in terms of timing. Both intra and interspecific variations are observed that hint at possible evolutionary processes: more overlap and communal chorusing in tight social groups, more alternation between distant neighbors, with sometimes both types of exchanges in the same species according to context. There are suggestions that temporal regulation would depend upon both development and social influences.

To date, there is a clear lack of targeted studies on particular animal models where all these facets could be investigated. Very few primate studies and almost no songbird study has considered the context of these different types of exchanges together with developmental issues, and even fewer are devoted to the cognitive (perceptual) processes involved. Comparative work is also often lacking, or draws on species other than those studied in terms of proximate factors. To test the possible social bases for the evolution of temporal aspects such as the turn-taking, we also need to study species from a common phylogenetic lineage, which differ in their social organization.

We will here try and tackle these questions on one songbird species, the European starling *Sturnus vulgaris*, well known for its vocal and social richness, and which has become one of the classical animal models for song studies (e.g., Eens, 1997; Hausberger, 1997). Comparative data from other Sturnid species are now available.

Testing Turn Taking in an Animal Model: The Starling

European starlings are highly gregarious birds that form breeding colonies of a few nests, which can be considered as the basic social unit, especially in sedentary populations (Clergeau, 1989). They forage in flocks from 10 to several hundred birds, and gather in the evening at roosts where several hundred to several thousand birds can be present (Feare, 1984). In all these contexts, song

is produced (Adret-Hausberger, 1982). The males spend every morning in their colony (all year round in sedentary populations, in Spring in migratory populations), they visit their nest and have vocal interactions with their colony neighbors, which are generally from 5 to 20 m away. The colony membership tends to be stable over time, although some birds may disappear and be replaced (Adret-Hausberger et al., 1990). Neighbors are therefore familiar. Males defend only the nest vicinity. Vocal interactions between neighbors involve particular vocal structures which are loud simple whistles that are produced with silent intervals between successive whistles, intervals where responses from other birds generally occur (Hausberger, 1991). As in all songbird species, starlings produce both calls and song. Calls are short and simple vocalizations produced in particular contexts for which an immediate function can be identified. Birds produce alarm calls, distress calls or flight calls, for example (Thorpe, 1961). Songs are more complex vocalizations whose functions are not so immediately obvious. Songs are produced in social contexts as well as breeding contexts. Starlings are able to produce two different categories of songs: whistles and warbling (Adret-Hausberger and Jenkins, 1988; Eens et al., 1989). These two categories of song are different in structure and in function as well as in their pattern of acquisition (George et al., 2010). Whistles are short, loud and stereotyped vocalizations that are produced in a discontinuous way. By contrast, warbling is characterized by its complexity and low intensity and consists of successions of motifs (a fixed combination of acoustic elements) produced in unbroken sequences for up to a minute (see also Chaiken et al., 1993).

Whereas whistles can be produced independently, warbling is often preceded by whistles and it then shows a clear organization based on repetition of motif types and an increase in tempo and frequency ending with clicks and followed by high-pitched trills (**Figure 1**). Warbling is not used in alternating vocal interactions and is mostly sung solo in the field. Playback experiments show that the birds react to whistled structures by replying vocally while they do not respond nor change their behavior when warbling (pers. obs). The developmental course of these two categories of songs is different (Poirier et al., 2004; Bertin et al., 2007). Warbling develops progressively from subsong in the course of the bird's first year of life, whereas whistles appear suddenly during the first winter around 9 month of age (Adret-Hausberger, 1989). Moreover, young birds raised without direct contact with adults will not develop whistles but will produce warbling song (Poirier et al., 2004; Bertin et al., 2007). Finally, neuroethological as well as functional magnetic resonance imaging (fMRI) studies performed on starlings revealed that these two distinct categories of song are not processed in the same way in the brain (George et al., 2008; De Groof et al., 2013).

Here we will focus on the singing style that emerges from the use of one or the other of these song categories, one discontinuous and enabling alternating interactions, the other continuous and hence not appropriate for turn-taking types of interactions.

One other interesting feature is that these two categories of songs have been found in other Sturnids such as the Indian hill mynah *Gracula religiosa* (Bertram, 1970) or the wattled starling *Creatophora cinerea* (Sontag, 1991) suggesting that comparative studies within this family of songbirds could be promising for understanding the evolutionary roots of the temporal regulation of vocal interactions.

In the following section, we describe a series of observations and experiments on the European starling, followed by field data on other sturnids, in order to examine the different facets of temporal regulation of interactions in one species in relation to the four questions of Tinbergen (1963): causation (why do these temporal features appear now and how are they processed?), ontogeny (how did they develop at the individual level?), function (what are their immediate functions?) and evolution (what adaptations led this species to develop these forms of interaction?).

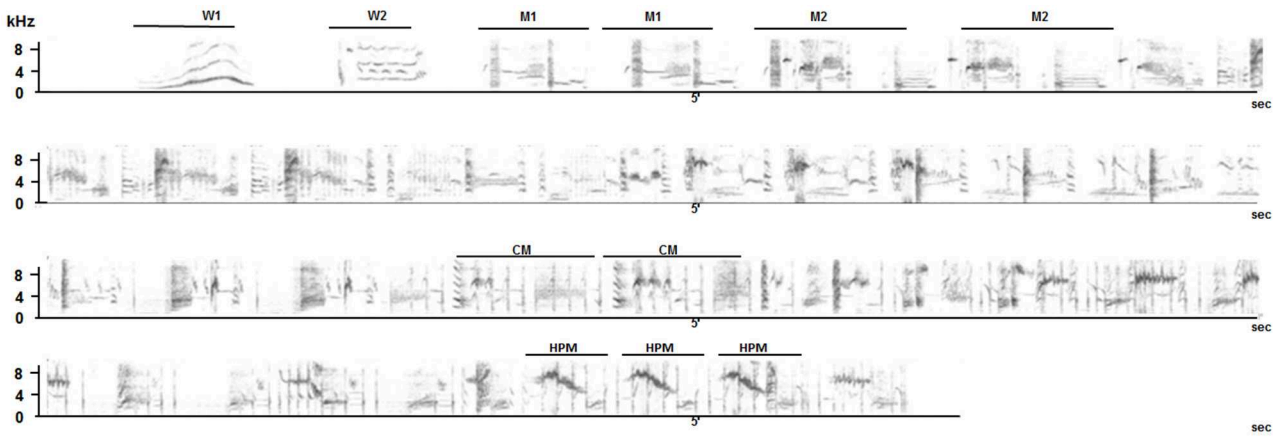


FIGURE 1 | Song sequence of a wild starling. The typical song sequence starts with two different types of whistles (W1 and W2). The warbling sequence starts with variable motif types (M1, M2 etc…) that can be repeated several times. Click motifs (CM) appear in the middle of the sequence. High pitched trill motifs (HPT) are characteristic of the end of the sequence.

Promoting or Not Turn-taking: Does the Social Situation Influence Temporal Features of Song?

Here we compared the singing style of European starlings living in colonies at different social densities.

Methods

Song recordings from 21 birds were re-analyzed in order to examine their singing styles (warbling/whistles). Data were available for four adult males recorded in isolation in captive conditions and 17 adult males recorded in the morning near their nests in the field. Recording sites and dates are shown in **Table 1**. More details about the recording conditions are given in the references mentioned. Additional aviary recordings when in a large mixed group were also available for the 4 isolated birds (Hausberger et al., 1995) (**Table 1**).

All field recordings were made in the morning during the first hours of daylight or during the two last hours when song is most frequent during spring in the breeding colony. Most birds were paired (most recordings are from sedentary populations). We recorded only adult males which had visited a nest, and were singing close to their nest. These recording sessions lasted one to several hours. The colony size was noted: a bird nesting singly or in colonies of 2, 3 up to 18 nests. Two nests were considered as belonging from different colonies when they were more than 200 m. apart (Hausberger and Guyomarc'h, 1981). Since colonies of 6–8 nests, 9–11 nests, 12 and 13 nest and 14–18 nests showed the same trends and the number of such colonies was low, we pooled the corresponding data. The captive males had been caught on Jersey Island. They were kept in sound proof chambers at day lengths corresponding to the natural photoperiod. Birds had water and food *ad libitum* (commercial pellets for turkeys, and apples). Recordings were made continuously for 4 consecutive days for each isolated bird. Recordings were made using different tape, or cassette- recorders and microphones (see references). Sound analyses were carried out on an Amiga microcomputer (Richard, 1991). We considered that different elements belonged to the same song bout when they were separated by less than 20 s. This was based on data on whistled sequences showing that successive whistles within a sequence can be separated by up to 12 s (Hausberger, 1991). A warbling sequence corresponds to a succession of elements separated by less than 1 s (Adret-Hausberger and Jenkins, 1988; Hausberger, 1997). Since different studies are summarized here, the recording times were different for the different birds and therefore the absolute number of bouts, warbling or whistle sequences could not be compared between birds. This study was carried out in accordance with the recommendations of European Communities guidelines (European Communities Council Directive of 24 November 1986 (86/609/EEC). The protocol was approved by the local Ethic Committee in Animal experiment of Rennes (CREA-07).

TABLE 1 | Recording sites, social conditions for all the birds.

Birds	Captive birds	Wild birds										
	CM1-CM4	WM1	WM2	WM3	WM4	WM5	WM6	WM7-WM8	WM9-WM11	WM12-WM13	WM14-WM17	
Social conditions	Isolation	1 nest	1 nest	1 nest	1 nest	2 nests	3 nests	3 nests	6 nests Colony 1	11 nests Colony 2	18 nests Colony 3	
Sites	Rennes (F)	Auckland (NZ)	Seewiesen (Germ)	Slimbridge (G.B)	Rennes (F)	Rennes (F)	Rennes (F)	Nouvoitou (F)	Rennes University Campus (F)			
Pairing Status	None	Unpaired	Paired	Paired (2 0)	Paired	Paired	Paired	Paired	Paired			
Dates recordings	January 1992	Spring 1987	Spring 1988	March 1990	March 1989	March 1992	March 1990	March 1979	March 1979 March 1982			
References	Hausberger et al., 1995	Adret-Hausberger and Jenkins, 1988	Unpubl.	Hausberger and Black, 1990	Unpubl.	Unpubl.	Henry et al., 1994	Hausberger and Guyomarc'h, 1981	Hausberger and Guyomarc'h, 1981; Adret-Hausberger, 1986			

Results

Individual adaptations to the social situation (Appendix 1 in Supplementary Material)

For four males, their song had been recorded both when in a mixed group in an outdoor aviary and while they were in isolation in sound proof chambers When isolated, these males

produced songs that included at least one warbling sequence whereas whistles were not always present. Almost all ($X = 92.53 \pm 7.89\%$) whistle sequences were followed by warbling, which was also generally preceded by a whistle (see Appendix 1 in Supplementary Material). The whistle sequences were composed of a very low number of whistles (mostly 1–3) (compare to Hausberger, 1991). The proportion of warbling and whistle sequences was similar for two birds but the two other birds showed a higher number of warbling sequences than whistle sequences. Thus, overall, warbling clearly predominated in this context, given also its longer time duration.

When the same birds were observed in a group, they showed a lower proportion of sequences including warbling ($Xi = 98.1 \pm 9.95$, $Xg = 65.42 \pm 11.35$, Fisher test for the 4 males, $p \leq 0.05$), a lower proportion of whistle sequences followed by warbling ($Xi = 92.53 \pm 7.89\%$, $Xg = 55.95 \pm 15.31$, Fisher test, $p \leq 0.05$ for three out of the fourmales) and also a lower proportion of whistles per sequence ($Xi = 1.98 \pm 0.48$, $Xg = 1.51 \pm 0.29$, *t*-test, M1 $p \leq 0.05$ for the four individuals) (see Appendix 1 in Supplementary Material).

Isolation vs. field recordings (Table 2 and Appendix 1 in Supplementary Material)

Compared to the songs of isolated birds, field recordings revealed a lower proportion of sequences including warbling (Mann Whitney, $n_1 = 4$, $n_2 = 17$, $U = 0$, $p < 0.002$), a lower proportion of whistled sequences followed by warbling ($U = 0$, $p < 0.002$) and a lower proportion of whistles followed by warbling ($U = 4$, $p < 0.05$) while the number of whistles per sequence was lower in isolation ($U = 1\, p < 0.002$). The proportion of whistle sequences compared to warbling sequences was overall much higher in the field as well as the ratio of the whistle sequences and warbling sequences ($U = 2$, $p < 0.001$) in both cases. Isolated birds and wild birds in all situations did produce the same proportion of song bouts including at least one whistle ($U = 32$, $p > 0.05$).

The importance of colony size (Table 2 and Figure 2)

Clear differences appeared in the singing style of birds according to colony size. As colony size increased, we found:

- a decrease in the proportion of bouts including warbling (Spearman test, $N = 17$, $r_s = -0.89$, $p < 0.0004$) in relation to colony size (Kruskall Wallis test, $H = 12.5$, $n_1 = 4$, $n_2 = 4$, $n_3 = 3$, $n_4 = 6$, $p < 0.0006$).
- a decrease in the proportion of whistle sequences followed by warbling ($r_s = -0.81$, $p = 0.001$) with differences according to colony size ($H = 9.8$, $p < 0.02$).
- a decrease in the proportion of whistles followed by warbling ($r_s = -0.89$, $p = 0.0002$) with differences according to colony size ($H = 12.9$, $p < 0.005$).
- an increase in the mean number of whistles per sequence ($r_s = 0.83$, $p = 0.001$) with differences according to colony size ($H = 10.5$, $p = 0.02$).
- an increase in the ratio of the number of whistle sequences to the number of warbling sequences ($r_s = 0.9$, $p = 0.003$) with differences according to colony size ($H = 13.1$, $p = 0.004$) (**Figure 3**).

TABLE 2 | Song characteristics of each individual.

Birds	WM1	WM2	WM3	WM4	WM5	WM6	WM7	WM8	WM9	WM10	WM11	WM12	WM13	WM14	WM15	WM16	WM17
Total number of song bouts	120	47	97	62	49	52	74	23	32	45	26	47	31	24	27	25	35
Total number of whistles	331	130	146	103	112	126	198	62	124	268	58	346	162	135	158	208	211
Proportion bouts including warbling % (Number)	67 (81)	77 (36)	60 (58)	84 (52)	55 (27)	40 (21)	61 (45)	57 (13)	53 (17)	40 (18)	58 (15)	53 (25)	17 (5)	8 (2)	7 (2)	12 (3)	29 (10)
Proportion bouts including whistles % (Number)	92 (110)	100 (47)	80 (78)	69 (43)	82 (40)	87 (45)	82 (61)	91 (21)	81 (26)	96 (43)	77 (20)	98 (46)	97 (30)	92 (22)	100 (27)	100 (25)	89 (31)
Proportion whistle sequences followed by warbling % (Number)	65 (71)	77 (36)	50 (39)	80 (33)	42 (18)	31 (14)	52 (32)	52 (11)	42 (11)	37 (16)	45 (9)	52 (24)	13 (4)	0 (0)	7 (2)	12 (3)	19 (6)
Number whistle per sequence X ± sd	3.1 ± 1.9	2.7 ± 1.8	2 ± 1.5	2.5 ± 1.8	2.4 ± 1.6	2.9 ± 1.7	3.2 ± 3.3	2.9 ± 2.9	4.8 ± 4.4	6.2 ± 4.6	2.9 ± 1.8	7.5 ± 5.4	5.4 ± 4.2	6.1 ± 6.0)	5.9 ± 3.7	8.3 ± 8.4	6.8 ± 6.2
Proportion of whistles followed by warbling %	21	28	27	32	16	11	16	18	9	6	18	7	2	0	1	1	3
Proportion warbling preceded by whistle	88	100	67	63	67	67	71	85	65	89	60	96	20	0	100	100	60
Number whistle sequence/warbling sequence	1.3	1.3	1.3	0.8	1.6	2.1	1.4	1.6	1.5	2.4	1.5	1.8	6	11	13	8.3	3.1

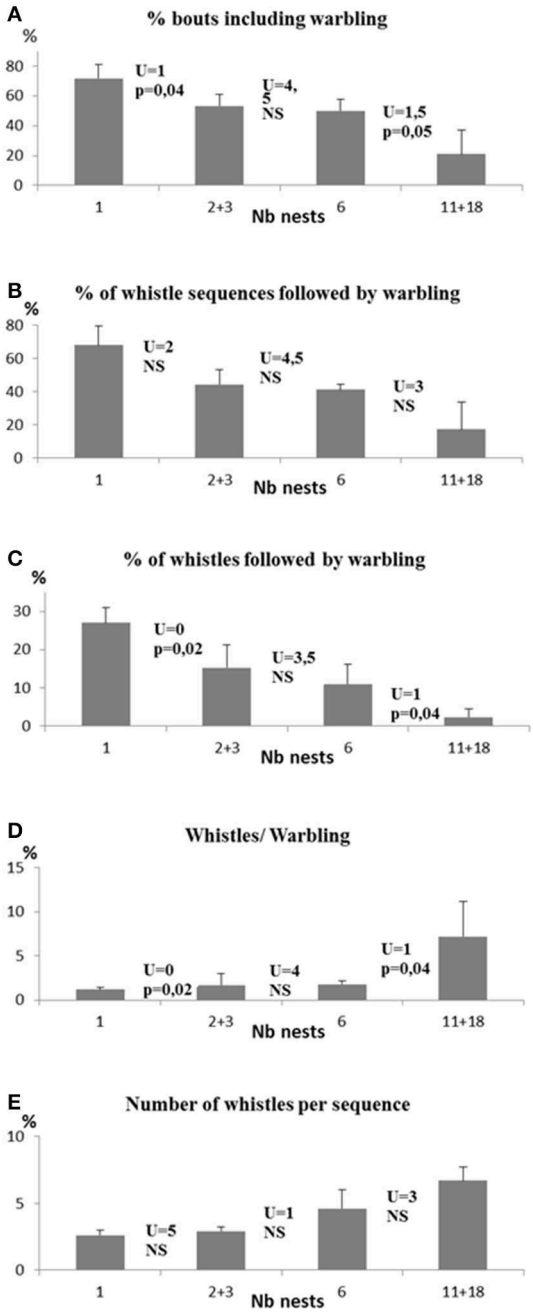


FIGURE 2 | Differences in song characteristics according to the size of the colony. (A) % of bouts including warbling; **(B)** % of whistle sequences followed by warbling; **(C)** % of whistles followed by warbling; **(D)** ratio Whistles/Warbling; **(E)** number of whistles per sequence.

While all these comparisons were verified when colony sizes were compared pairwise, the colony sizes 2/3 and 6 nests did not show significant differences in most of the above mentioned comparisons. Additional differences were observed between colony size 1 and 11/18 for the percentage of whistles following by warbling and the number of whistles per sequence (Mann Whitney $U = 1, P < 0.003$ and $U = 0, p < 0.005$ respectively) (see **Figure 2**) suggesting that the overall trend is accentuated in extreme social situations.

Conclusion

Singing style is clearly influenced by the social situation in male starlings. The more birds there are around them, especially in the breeding context, the more they favor the production of discontinuous songs, which is a prerequisite for alternating vocal exchanges. In large colonies, male starling song showed a high proportion of whistles, leaving much opportunity for interactions and transfer of information between neighboring males (**Figure 4**). Data from breeding sites where the birds nested singly were similar to those obtained in isolated captive birds, revealing that it is more the presence of potential vocal partners than the presence of another bird (mate) that influences the choice of a singing style. Comparison of the same birds in different contexts revealed that there is an individual capacity to adjust the singing style to the social situation.

Are There Temporal Regularities in Starling Vocal Interactions?

Here the immediate responses of male starlings in terms of temporal opportunities for response in the presence of another individual and its interactional status were observed in spontaneous interactions.

Methods

Seven male starlings were observed in the same breeding colony (4 in 2002, 3 in 2003) between March 17th and 27th 2002 and between March 17th and April 15th in 2003 from 7 am (sunrise) to 11 a.m. All were paired at that time of the year. The colony was composed of 5–6 pairs. This study was carried out in accordance with the recommendations of European Communities guidelines (European Communities Council Directive of 24 November 1986 (86/609/EEC). The protocol was approved by the local Ethic Committee in Animal experiment of Rennes (CREA-07).

Each full whistle sequence of the focal bird was recorded until it started warbling or left. Two contexts of singing were considered: 1- singing alone with no congener present in the vicinity and 2- singing with another male silent or singing in the vicinity.

Song recordings were made using a Sony TC D5 cassette recorder and a Sennheiser directional microphone (MZA 14 P48) in 2002, or a Sony microphone (EMC 144) fixed on a polyester parabola in 2003. Vocalizations were analyzed using a computer (Unix Silicon Graphics Ind), and a custom-designed sound analysis software (ANA, Richard, 1991).

Results

We plotted the intervals between successive whistles produced by two different birds (**Figure 5**). More than half of the whistles (56.4%) were produced within 2 s. We thus considered that two whistles separated by 2 s or less belonged to a single vocal interaction (see also Adret-Hausberger, 1982; Miller et al., 2004). Eight hundred and thirty five whistles were recorded in total.

The intra-individual interwhistle interval (IWI) clearly increased when another starling was singing nearby ($Xa = 4.7 \pm$

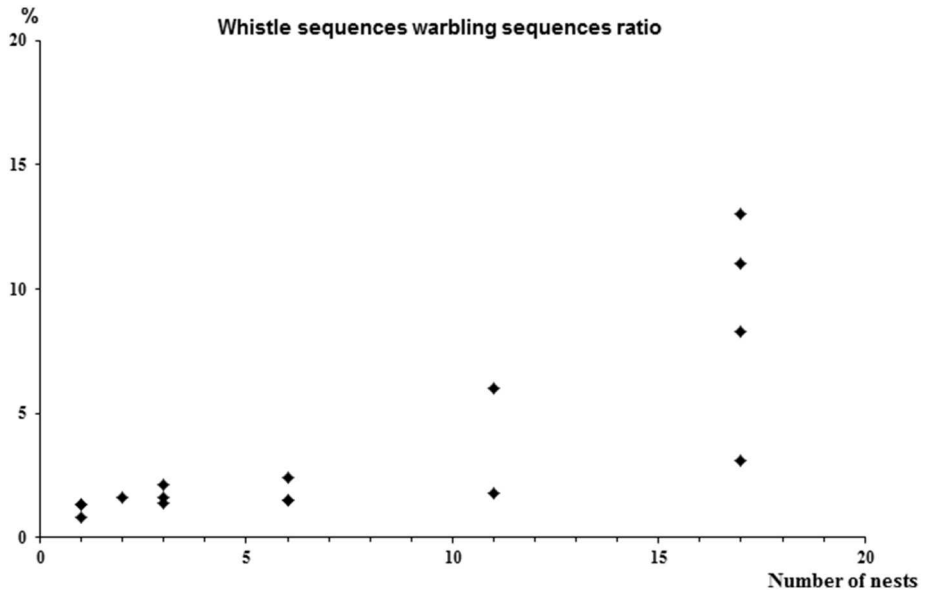


FIGURE 3 | Correlation between the number of whistle sequences and the number of warbling sequences ($r_S = 0.9$, $p = 0.003$).

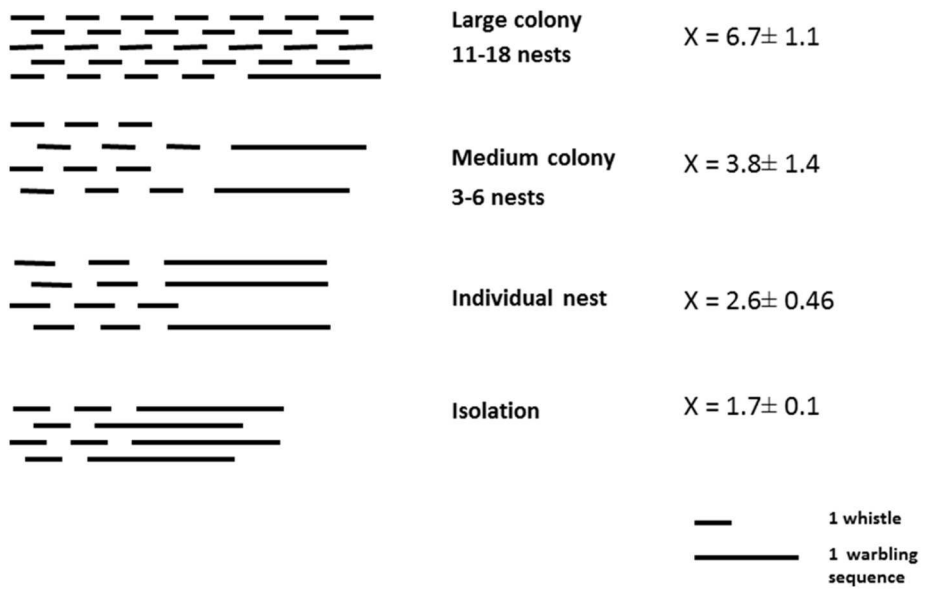


FIGURE 4 | Song style of birds belonging to colonies of different size. Although the birds were recorded in very different conditions, a clear trend appeared toward an increase in whistling (hence discontinuous songs) and a decrease of warbling (hence continuous song) with increasing colony size (= number of neighbors) (From Hausberger, 1997).

1.3 s, Xns = 11 ± 1.6 s Wilcoxon, $N = 7$, $T = 0$, $p < 0.02$ (**Figure 6A**). Indeed, five of the seven males doubled this interval and one quadrupled it.

For four of the birds, we recorded sessions when the neighbor was silent: clear differences appeared again: the IWI did not differ significantly between the solitary situation and the "silent neighbour" situation ($Xa = 4.3 \pm 1$ s, Xnst = 5.11 ± 0.65 s, Mann Whitney, $n_1 = n_2 = 4$, $U = 5$, $p > 0.5$ while the IWI in the "singing neighbour situation" differed from both (Xns = 10.68 ± 0.93 s, MW; alone/neighbor singing, $U = 0$, $p \leq 0.05$ in both cases) (**Figure 6B**). Only 133 instances of overlapping (second emitter started before the end of the whistle) were observed, but in 83% of the cases they were associated with the end of vocal exchanges (first emitter became silent or flew away), which is more than expected by chance ($X^2 = 63.11$, $df = 1$, $P \leq 0.001$) (**Figure 6C**).

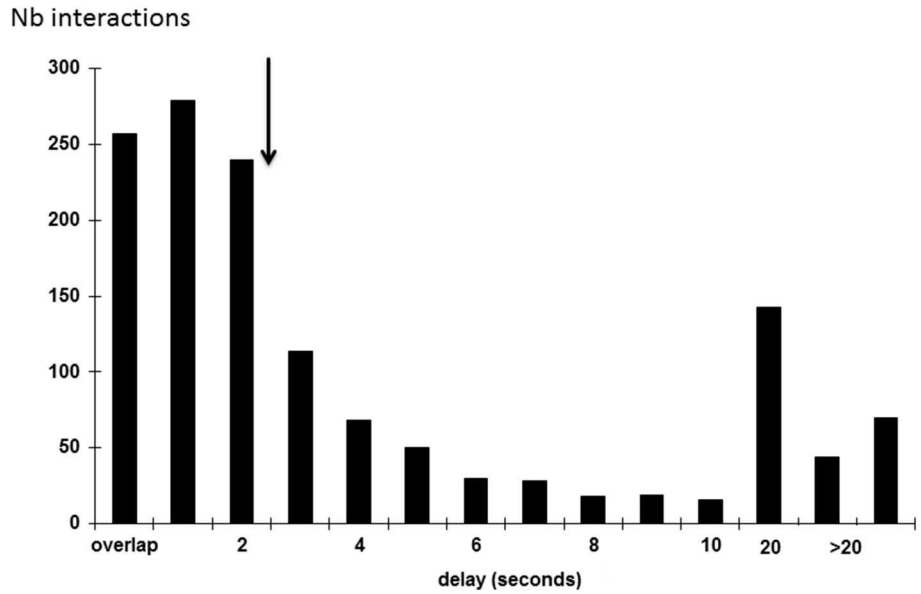


FIGURE 5 | Interval separating two successive whistles produced by two different individuals (overlap: when two whistles overlap). The arrow indicated a break in the interactions after a 2 s delay.

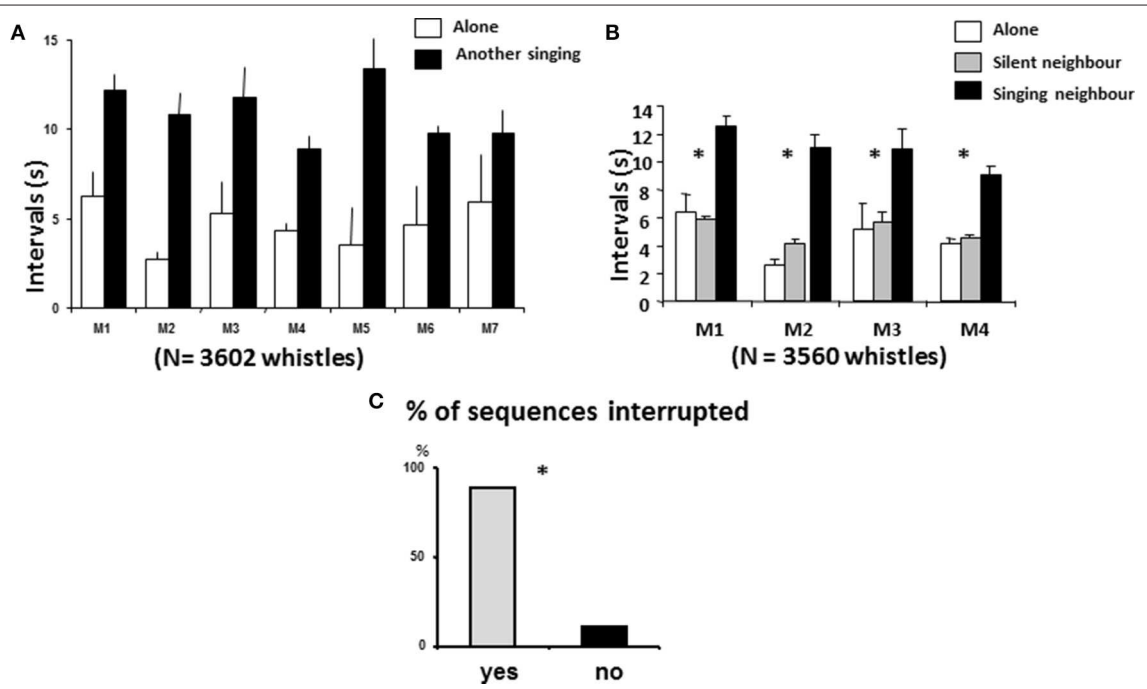


FIGURE 6 | Song behavior according to immediate context. (A) Males increased their interval duration when another bird was singing (Wilcoxon, $N = 7$, $T = 0$, $p < 0.02$). **(B)** Birds did not change their interval duration when another bird was present but silent and increased their interval duration when the other bird was singing. **(C)** Most of the birds interrupted the vocal interaction in case of overlap (* : $X^2 = 2256$, $df = 1$, $p \leq 0.05$).

Conclusion

It appeared that starlings take into account the social context when they are singing. By increasing interval duration between two whistles, starlings clearly leave space for other birds to reply and therefore make turn-taking possible. Another element that showed evidence of "conversation rules" in the starling was

a response overlap between whistles from two males, which appears here as "breaking the rule" and led to the end of the exchange.

Developmental Issues: How Do Young Birds Acquire an Appropriate Singing Style?

The impact of developmental conditions, both on the sensory and social levels, has been tested through a series of experiments.

Normal Development

Young starlings like other songbirds develop their songs slowly with distinct stages, starting with "subsong," at the age of about 3 months: a long, continuous, disorganized vocalization where the young bird is just practicing, and then a plastic stage where elements of the future song appear progressively. It has been suggested that subsong and plastic song are analogous to infant babbling (e.g., Marler, 1970).

Also like other songbirds, starlings need to hear adult song in order to develop normal songs (e.g., Chaiken et al., 1993). Little attention has been paid in the developmental studies of starling song or even other songbirds to how developmental stages might affect turn-taking responses.

Field observations are almost impossible as the young birds disperse and become nomadic after fledging (Feare, 1984), thus only some data from captive birds are available (they are also difficult to breed in captivity). Monitoring nine young males from birth to adulthood in an aviary where they were kept with their parents confirmed anecdotal reports from the field in terms of the timing of subsong and plastic song but also revealed that the first whistles (hence discontinuous songs) were produced in November, at the age of 7 months. Until then, only continuous song was produced although the plastic song starts showing some disruption (**Figure 7**).

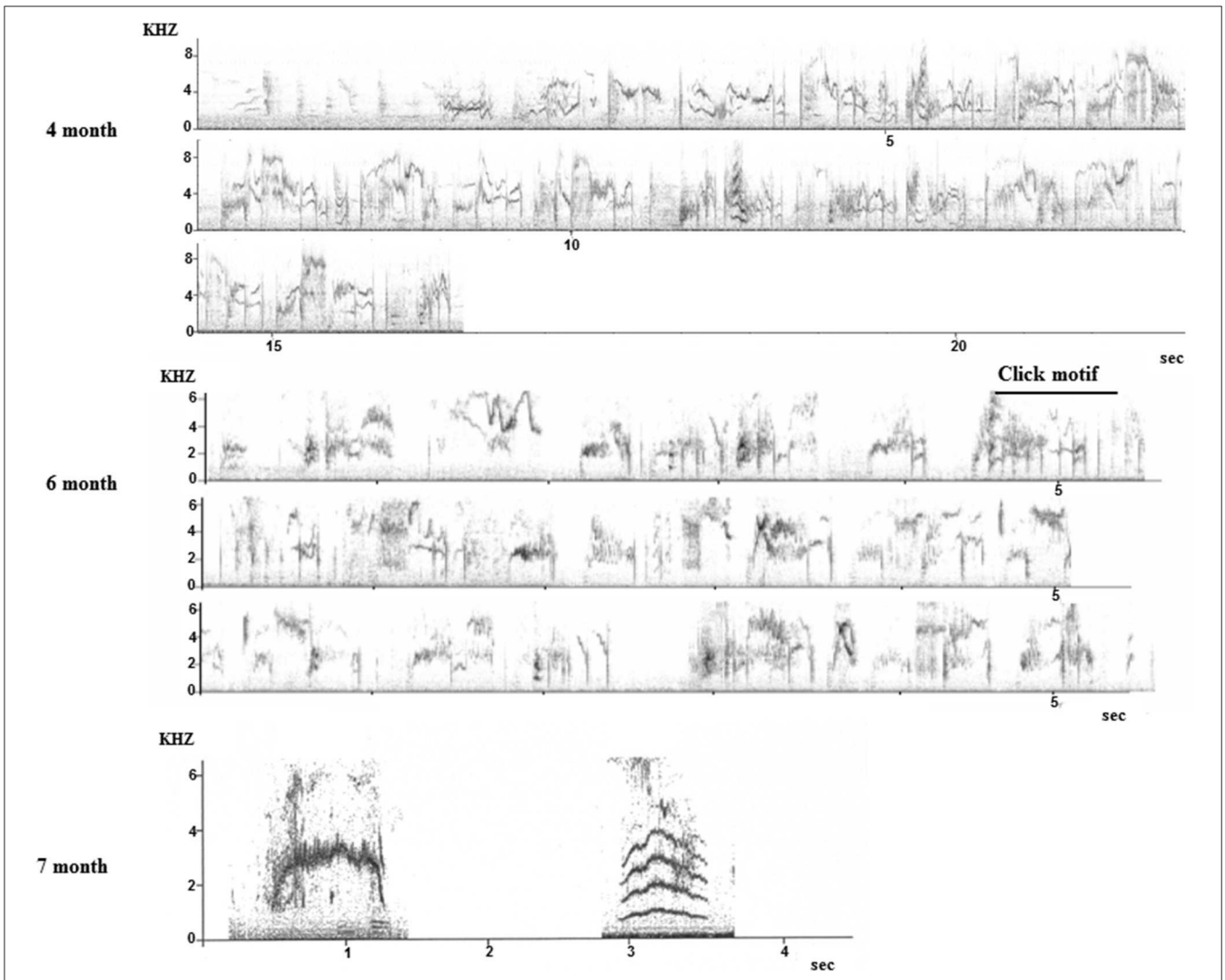


FIGURE 7 | Sonograms of song produced by young starlings during the first year. Top: subsong produced during the first summer (4 months old). **Middle**, sequence produced at 6 months: click motifs are recognizable. **Bottom**: two whistles recorded at 7 months old.

Another pilot study on one young male raised without adults but able to hear adult birds showed the same trend with the first whistles appearing at the age of 9 months (Adret-Hausberger, 1989). In all cases, the whistles appeared suddenly and quite independently from subsong that seemed to develop progressively into the adult warbling. We noted that the first click motifs appeared in the subsong at 6 months and subsong sequences progressively showed more adult-like organization (**Figure 7**). It has been proposed in starlings as in other songbirds that warbling types of songs (long and continuous) could be an adult form of subsong (Adret-Hausberger, 1989).

Disturbed Ontogeny: The Importance of Adults

Sensory and physical deprivation of experience with adults

Existing data on starlings raised without exposure to adult song or contact with adult birds were reanalyzed in order to extract information on their singing style. Four male starlings were taken as nestlings (2–5 days old) and hand raised without any contact with adults. They were kept respectively in groups of inexperienced animals: 1 male with 4 females of the same age (May 1993), 2 males and one female of the same age (May 1992) and one male amongst other clutches of 19 other males and females (May 1981). Their song was recorded when adult at 1 year old. This study was carried out in accordance with the recommendations of European Communities guidelines (European Communities Council Directive of 24 November 1986 (86/609/EEC). The protocol was approved by the local Ethic Committee in Animal experiment of Rennes (CREA-07).

Forty-five to 123 song sequences could be recorded from each individual. None of them ever produced a whistle. They all sang a continuous song that showed some similarities to a "normal adult warbling" especially in its continuous type of structuring (**Figure 7**). While separate motifs appear, the intermotif intervals were, as in a normal adult song (e.g., Eens et al., 1989) too short to permit a non-overlapping response from another bird ($X = 0.19 \pm 0.18$ to 0.59 ± 0.25 s) (**Figure 8**).

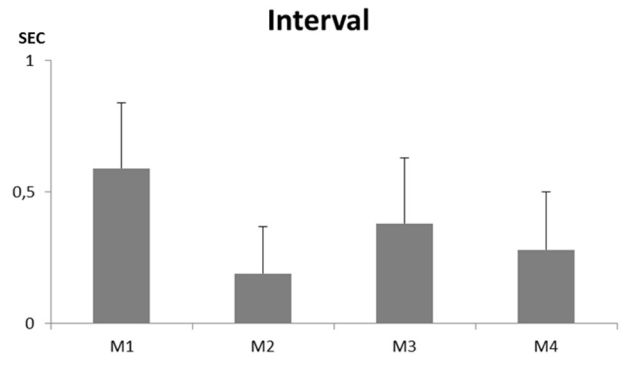


FIGURE 8 | Intermotif intervals recorded for four young male starlings raised without exposure to adult song or contact with adult birds.

Varying the type of contact with adults (Poirier et al., 2004)

This experiment involved 26 young starlings taken from the nest in April 1998 when 2–5 days old and then hand raised for 2 months. In June 1998, they were placed in one of three situations: eleven (5 males) were placed in groups of 3 or 4 in three aviaries together with wild caught adult males in indoor aviaries; 6 (4 males) were kept in isolation and 6 (4 males) in pairs of inexperienced birds in sound proof chambers fitted with loudspeakers that transmitted the sounds from the aviary room (**Figure 9**). The isolated and pair raised animals could thus continuously hear the vocal interactions that occurred in the aviaries. This study was carried out in accordance with the recommendations of European Communities guidelines (European Communities Council Directive of 24 November 1986 (86/609/EEC). The protocol was approved by the local Ethic Committee in Animal experiment of Rennes (CREA-07).

Their song was recorded when they were adults in 1999 after 1 year spent in this situation. The recordings show that all 5 males raised in direct contact with an adult produced whistles while of the others, only the two isolated males did so. They were also those whose output showed the closest resemblance to adult song. The pair-raised animals did not produce any discontinuous song, hence separate whistles. They did produce some whistled notes but these were included in a warbling sequence with no time interval. They also had a very variable warbling song much like juvenile subsong.

Because they had no other sensory stimulation, isolated birds paid more attention to the adult song heard through the loudspeaker and hence developed some discontinuous songs (Poirier et al., 2004). It remains to be established if they will use them in an appropriate way. Further studies seem to indicate that the absence of adult contact during development prevents the development of a normal singing style and proper use of song types (George et al., 2010).

Appropriate social contact during development is thus necessary and crucial in order to produce songs that enable an alternating communication pattern. The birds raised in the aviaries with one adult model nevertheless still showed some abnormalities that pose questions concerning the importance of the adult-young ratio. They formed mostly small same-sex age groups that sang together with mainly overlapping vocal interactions.

Testing the impact of the adult-young ratio (Bertin et al., 2007)

Twenty male starling nestlings (6–8 days old) were taken from the nest in April 2002. They were hand raised and were kept as a mixed social group with 27 peer females until the age of 2 months with no contact with any adult. In June, they were allocated to three different rearing conditions: (1) dyadic: one adult-one young, (2) group tutored: 7 young and 2 adults, (3) group: 5 young birds together. The groups could hear but not see the other animals (which were housed in the same room), thus providing a similar auditory environment. This study was carried out in accordance with the recommendations of European Communities guidelines (European Communities

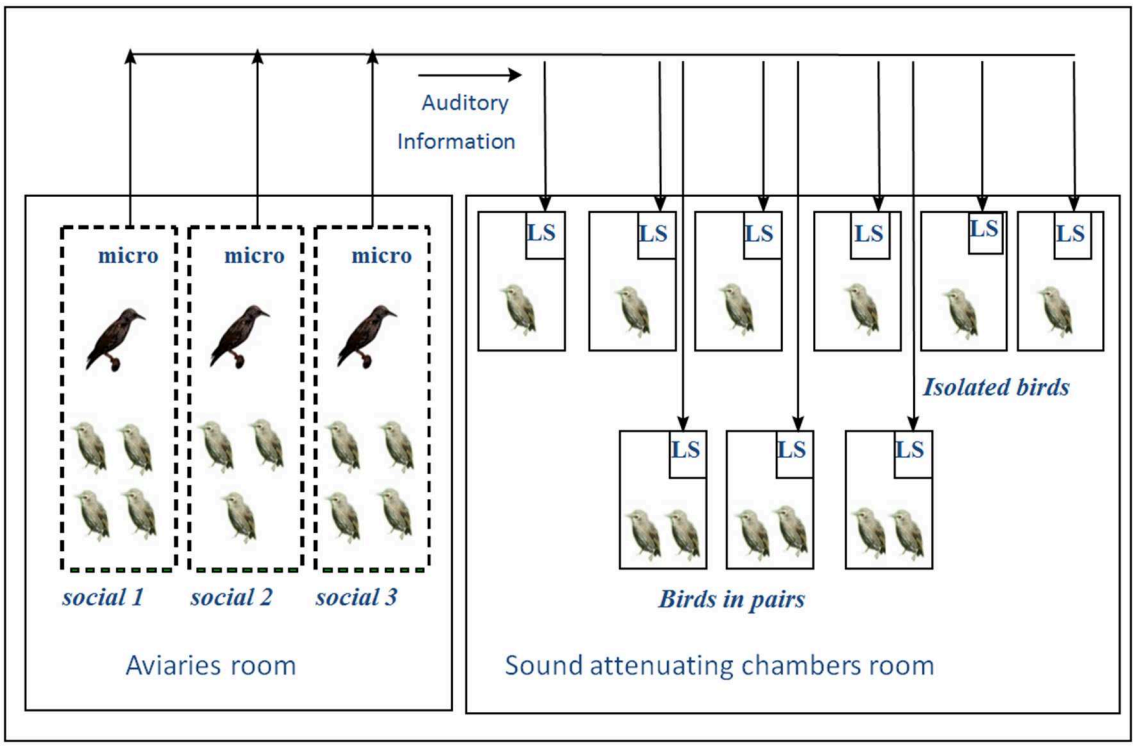


FIGURE 9 | Experimental set-up: In the aviary room, young birds housed in the three different aviaries could hear and see each other. In the soundproof chambers, birds were housed individually or in pairs. All the experimental birds received the same auditory exposure. Birds in the soundproof chambers could hear all the songs from the birds in the aviaries via microphones and loudspeakers. (M, wild males; m, experimental males; f, experimental females; LS, loudspeaker; micro, microphone; →, direction of auditory information transfer (from Poirier et al., 2004).

Council Directive of 24 November 1986 (86/609/EEC). The protocol was approved by the local Ethic Committee in Animal experiment of Rennes (CREA-07).

The song of the experimental animals was recorded and analyzed when they were 1 year old. The results show that the repertoire of whistles followed a gradient with fewer whistle types in the group tutored than in the dyadic situation and almost no whistles produced by the peer-only group (only 1 whistle type in two of them).

When still in their developmental setting, both groups (group tutored or not tutored) sang more (in time) than the animals placed in the dyadic situation, but since their song repertoire was mostly or only composed of warbling, they kept singing together, overlapping without any temporal organization.

Brain Mechanisms and Plasticity: The Processing of Song Categories and the Effect of Experience

It was hypothesized that if the two different song categories (discontinuous/continuous) had a different functional significance and as shown above, different developmental trajectories, the brain processes involved should be to some extent different. The following studies were carried out in accordance with the recommendations of European Communities guidelines (European Communities Council Directive of 24 November 1986 (86/609/EEC). All the protocols were approved by the local Ethic Committee in Animal experiment of Rennes (CREA-07).

Song Processing in Wild Caught Adults (George et al., 2004, 2008)

In a series of experiments on the processing of starling song in the brain, we tested the electrophysiological responses of field L (primary auditory area) and NCM (secondary auditory area) neurons of awake restrained adult (wild caught) starlings while they were exposed to a variety of species specific sounds (whistles, warbling elements) and artificial sounds (white noise, pure tones). Using a systematic approach to record neuronal activity (George et al., 2003), we were able to record the activity of almost 3000 neurons in the Field L and 2000 neurons in the NCM from 6 individuals each time.

It appeared that the distribution and level of response respectively varied according to the song category. There was lateralization of song processing so that in Field L, the whistles were processed more in the right hemisphere while the warbling was processed mostly in the left hemisphere, revealing a differential processing of these two categories of songs. In the NCM, which as a secondary area, processes more complex associative information (e.g., Chew et al., 1996), it appeared that most neurons responded first of all to songs bearing individual information, but both the proportion of responsive neuronal

sites and the magnitude of the neuronal responses differed according to the functional song classes. A gradient of response was observed from the class 1 whistle (eliciting the lowest level of responses), to the class 2 whistles and then warbling which clearly triggered more responses than the two classes of whistles.

Conclusion

Since the brain processes functional categories of songs differently and at different levels, it may trigger appropriate vocal production and enable the bird, when hearing one song category, to rapidly "decide whether or not" to reply.

How Can Social Experience During Development Affect Brain Processing of Song Categories? (Cousillas et al., 2004, 2006; George et al., 2010)

Responses of field L neurons of adult starlings raised without adults (no sensory contact) using the same procedure as above has revealed that the whole area (Field L) lacks the typical spatial organization of normal adults and also the typical neuronal selectivity toward specific song elements (Cousillas et al., 2004).

Social experience *per se* can evidently have as much influence on the development of the primary auditory area as the sensory experience in the experiment by Poirier et al. (2004). Thus, both the birds raised in pairs or solitarily showed as many abnormalities (lack of neuronal selectivity) as the sensory deprived birds. The lack of contact with adults was obviously sufficient to prevent proper development. Another intriguing finding was that even the birds raised in a group with one adult showed deficiencies, which seems to reflect their lack of social bonding with the adult (Cousillas et al., 2008).

Similar findings were obtained at the NCM level: 10 young birds were taken from the nest, hand raised, and then placed in a large outdoor aviary where they could hear wild adults but had no direct contact with any adult. Four months later they were transferred as a group to an indoor aviary with no auditory nor direct contact with adults for 12 months. These birds, when adult, had a fairly normal song repertoire including whistled and warbling structures. However, they did not produce sequences of whistles as "normal" starlings do (Hausberger, 1991), and placed them within warbling sequences which made them inappropriate for alternating vocal interactions (**Figure 10**). Interestingly, the electrophysiological recordings of the NCM neurons showed a clear deficiency in processing song categories (George et al., 2010). The lack of direct experience with adults despite a rich auditory experience therefore induced a singing style that did not promote alternation in vocal interactions despite the production of appropriate structures. Since brain processes devoted to song categorization were clearly affected, the birds probably could not recognize appropriate times for replying.

Conclusion

Social bonding and hence selective attention may be a key factor in developing the necessary brain processes and therefore the ability to communicate in an appropriate way.

Turn-taking as a Social Adaptation: An Evolutionary Process?

In the Eastern Cape in South Africa, four species of starlings with different social systems offered an opportunity to test the hypothesis that the temporal regulation of vocal interactions would reflect their social organization. The red-winged starling *Onychognathus morio*, pale-winged starling *Onychognathus nabouroup*, African pied starling *Spreo bicolor* and Cape glossy starling *Lamprotornis nitens* are widely sympatric in the region, but range from solitary pairs through colonial groups to

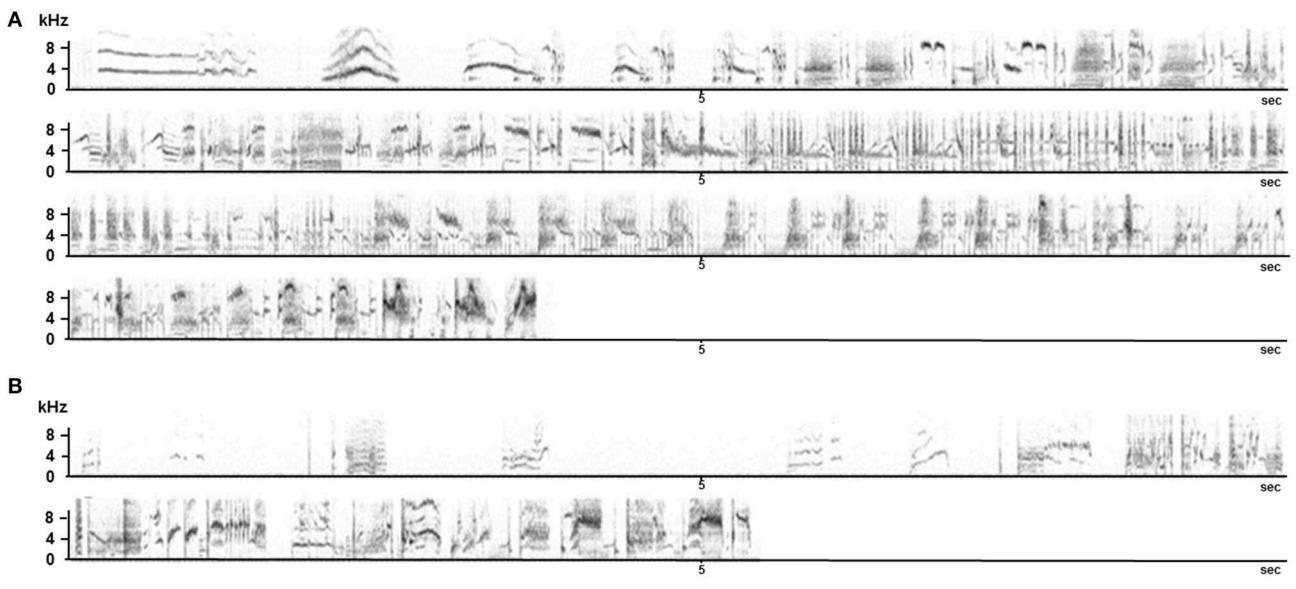


FIGURE 10 | Song sequences produced by an adult (A) and by a 2 year old birds that did not receive adult tutoring (B). Recordings were made at the same time of year.

communally-breeding species (Feare and Craig, 1999; Craig and Feare, 2009):

- A territorial species: *Onychognathus morio*

The red-winged starling is a sedentary species. Monogamous pairs remain together for at least three successive seasons and are associated throughout the year (Rowan, 1955; Craig et al., 1991). During the breeding season (October—March), pairs are restricted to their breeding territory (approximately 200 m^2) and very rarely join the flocks of non-breeding birds. Breeding pairs are extremely territorial and intraspecific aggression is very common. During the non-breeding season, both pairs and non-reproductive birds gather in flocks of varying sizes and spend the night together in large roosts (Craig and Feare, 2009).

- A colonial species: *Onychognathus nabouroup*

Monogamous pair bonds of the pale-winged starling are maintained throughout the year, and the birds apparently remain together for several seasons. At the beginning of the breeding season, male and female defend a small area around the nest, but pale-winged starlings seem clearly less aggressive than red-winged starlings. Throughout the year, birds roost in small flocks in groups on cliffs, with breeding pairs generally roosting at their nest site (Craig et al., 1991).

- A "familial" species: *Lamprotornis nitens*

Cape glossy starlings, a mainly sedentary species, breed in small family groups. Several monogamous couples nest in a same site (September—February). Nests may be in tree holes or other structures, and the same site is often re-used in successive years. According to Craig (1983) and Craig and Feare (2009), up to three birds, mostly young non-reproductive birds, help pairs to care for nestlings. During the non-breeding season, birds may gather in larger flocks of 10-20 birds to forage and share a regular roost site.

- A "communal" species: *Spreo bicolor*

Throughout the year, African pied starlings live in flocks of 15–25 individuals. Stable monogamous pairs re-use the same nest sites in successive breeding seasons (September—January). Several nests can be found close to each other in burrows or holes. During the breeding season, up to seven helpers can feed the young with the parents, and helpers may feed young at three different nests during a single breeding season (Craig, 1987). During the non-breeding season, pied starling groups may be nomadic and join other groups at communal roost sites (Craig and Feare, 2009).

Methods

This study was conducted in the Eastern Cape region in South Africa where the four species occur, often at the same sites. Songs have been recorded since 2003, mainly during the breeding season. This study was carried out in accordance with the recommendation of European Communities guidelines (European Communities Council Directive of 24 November 1986 (86/609/EEC). The protocol was approved by the local Ethic Committee in Animal experiment of Rennes (CREA-07).

Red-winged starling vocalizations were recorded primarily on Rhodes University campus in Grahamstown and in the vicinity (2003–2004). Pale-winged starling vocalizations were recorded at one site: Graaff-Reinet (2005). Pied starling vocalizations were recorded at three sites: Table Farm, Queenstown and Graaff-Reinet (2003 to 2005 and 2008). Finally, glossy starling songs were recorded at five sites: Thomas Baines Nature Reserve, Table Farm, Salem, Kariega Private Game Reserve (all in the Grahamstown area) and Queenstown (2003–2004).

From 2003 to 2005, a Sony TC-D5 Pro II tape recorder and a micro-directional microphone Sennheiser MKH 70 P48 were used to record vocalizations in the field. After 2005, we used a digital recorder Marantz PMD 660 and a directional microphone Sennheiser MKH 416 P48 (recordings made in 44.1 kHz/16 bits).

Most recordings were obtained in the morning (6–10 a.m.), and in the hottest hours of the day (12 a.m.–15 p.m.), corresponding to the peaks of activity of the studied birds (Feare and Craig, 1999). According to Fry et al. (2000), both sexes sing in all four species, despite the fact that, except for *O. morio*, males and females are not distinguishable. Vocalizations were analyzed using homemade software for song analyses (ANA, Richard, 1991). The amount of song recorded is summarized in **Table 3**.

Here we focused our analyses on the temporal aspects of songs. Indeed, most studies on interspecific comparisons of vocalizations have focused on quantitative aspects, such as the repertoire size (Catchpole, 1980; Kroodsma, 1977; MacComb and Semple, 2005). Whereas temporal aspects of vocal signals or vocal interactions have so far been little studied, they nonetheless could provide a wealth of information regarding the influence of social life on the evolution of vocal communication. We predicted that social life, in terms of the number of social partners or distance between partners for example, would affect the temporal structure of song.

We first estimated the proportion of discontinuous/continuous songs. Two categories of songs could be distinguished: *discontinuous songs*, corresponding to unitary notes or short motifs (a fixed combination of acoustic elements) produced at discrete intervals, and *continuous songs* in which long sequences are produced, with less than 0.5 s interval between two successive motifs.

For each species, we measured: 1- sequence duration, 2- intervals between two successive sequences or two successive discontinuous motifs, 3- the motif duration, 4- the number of motifs per sequence, 5- intervals between two successive motifs within a sequence.

Results

The four species showed clear differences in the temporal organization of their song. Considering the proportion of continuous and discontinuous songs, a gradient was observed from *O. morio,* that produced only single song elements (categorized as "whistles") to *S. bicolor* that produced only long phrases of continuous song (categorized as "warbling"). *O. nabouroup* and *L. nitens* appeared intermediate, producing both categories of songs (**Table 3**).

Interestingly, this gradient corresponded to the increase in the complexity of social life (**Figure 11**): the more the species showed

TABLE 3 | Song recordings for the four South African starling species and their song characteristics: temporal features (durations in seconds, mean ± SD); proportions of discontinuous songs (DS) and continuous songs (CS).

	O. morio	*O. nabouroup*		*L. nitens*		*S. bicolor*
Number of individuals	45	9		30		16
Total time of analyzed song (min)	6000	41		184		70
Total number of motifs analyzed	4500	1021		11,000		206
DS and CS song proportion (%)	**DS**	**DS**	**DC**	**DS**	**CS**	**CS**
	100	16.54	83.46	5.43	94.57	100
Motif duration	0.76 ± 0.23	0.15 ± 0.001	0.3 ± 0	0.32 ± 0.06	0.29 ± 0.04	0.17 ± 0
Phrase duration	0.76 ± 0.23	0.15 ± 0.001	1.79 ± 0.56	0.7 ± 0	2.91 ± 4.29	3.19 ± 2.07
Number of motifs per sequence	1 ± 0	1 ± 0	5.67 ± 1.91	2 ± 0	6.92 ± 3.12	11.78 ± 7.89
Duration between motifs	>1	>1	0.08 ± 0.04	0 ± 0	0.2 ± 0.04	0.13 ± 0.03
Duration between sequences	8.96 ± 4.58	2.99 ± 1.49	7.19 ± 3.68	4.07 ± 4	2.28 ± 1.20	2.06 ± 0.8

a complex and especially family type of social organization (in terms of number of congeners and nest proximity), the more their songs were produced in a continuous manner. In the same way, for species that produced continuous song, the phrase durations and the number of motifs per phrase increased following the same gradient (ANOVA, $F = 5.51$, $df = 2$, $p < 0.0001$; $F = 89.82$, $df = 2$, $p < 0.0001$ respectively).

On the other hand, the motif durations as well as the phrase intervals decreased following this "social" gradient (ANOVA, $F = 11891$, $df = 2$, $p < 0.0001$ $F = 442$, $df = 2$, $p < 0.0001$ respectively).

Song overlap was never observed in *O. morio*. On the contrary, in *L. nitens* and *S. bicolor*, song overlap was very common and we frequently recorded choruses of birds living in the same group (**Figure 12**). Both alternating and overlapping song interactions are also regularly observed in *O. nabouroup*.

Conclusion

The data presented here on one animal model reveal the interest of focusing on one question (here the temporal features of song that may or may not lead to alternating vocal interactions) and examining the different facets of the question. To the question: do European starlings show turn-taking in their vocal exchanges between males?, we can, from both observations and experiments, provide some answers: (1) they do favor alternation over overlapping, in particular through an immediate adaptation of the singer to the mere presence of another singing individual, but also according to the social situation and social density; (2) overlap does indeed disrupt the exchanges; (3) as in humans, there is an influence of context: alternation predominates in the usual interactions between males but chorusing can occur in more communal and intense social contexts (e.g., roosts), (4) the capacity for alternation develops during ontogenesis and social deprivation during development results in the inability to sing in a manner that favors "turn-taking" Social influences during development may directly affect the development of the brain processes devoted to song categorization.

While "turn-taking" is favored in the distant social interactions between males, more continuous song is produced in proximate interactions such as male-female interactions, or exchanges between close social partners (Hausberger et al., 1995). It has been proposed that warbling could play some stimulating role on the physiology of the listeners but also on the emitters too as found in budgerigars by Brockway (1969) and Adret-Hausberger and Jenkins (1988). Warbling is often associated with excitation behaviors such as visual displays and the production of high pitched trills, especially in the breeding season (Verheyen, 1980). As mentioned earlier, when producing warbling, male starlings seem to be "unaware" of the stimulation of their environment. Fundamentally, male starlings show movements of the head, typical of observation, during the silent interval between successive whistles and an erect posture while they are more in an oblique posture, with or without wing displays and a low reactivity while warbling.

In humans, it has been proposed that "attention is an intrinsic motivation for all utterances in a conversation, independent of the other possible motivation…" (Sacks et al., 1974). Excitation may lead to more overlap.

Interestingly, the comparative study of African starlings reflects these findings: the more communal the species, the more song overlap and choruses appear during close-range interactions, and the more continuous the song. The more territorial and long distant interactive a species is, the more alternation there is, hence the more discontinuous the song structures are. Some species like the European starling and the pale-winged starling show both song styles, reflecting the different contexts of interaction. Other species may also show this relationship between the temporal features of an interaction and the arousal states of the interactants: in barnacle geese *Branta leucopsis* triumph ceremonies, females that "encourage and support" their mate in the interaction will first alternate calling but with an increasing tempo and then overlap and chorus as excitation increases (Hausberger and Black, 1990) while those that do not support their partner (older pairs) produce other soft types of calls without any temporal synchronization (Bigot et al., 1995). According to Hauser (1992), the timing of calling

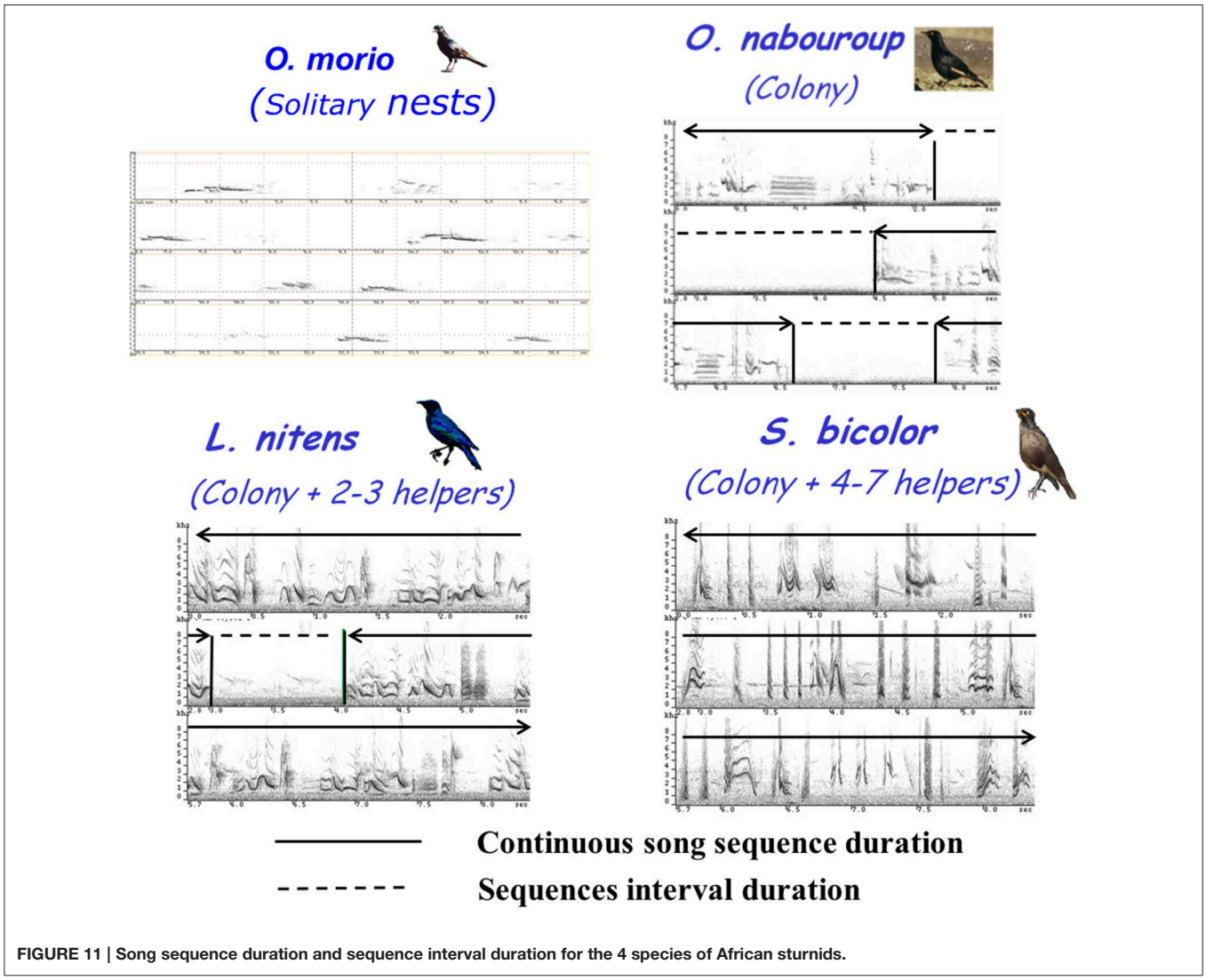


FIGURE 11 | Song sequence duration and sequence interval duration for the 4 species of African sturnids.

in macaques may be altered in such a way that it is used by individuals to manipulate or facilitate social relationships.

One may speculate that the need for mutual intelligibility and information seeking but also the need for giving and receiving attention, a potential mediator of social bonding (Fedurek et al., 2013) may have constituted the basis for the evolution of turn-taking. Humans too may produce choirs that are perceived as a communal display rather than an interaction between individuals.

In the Dogons, as mentioned earlier, observing rules in language coincides with law and order in the society (Calame-Griaule, 1965). It is true too that spacing of the vocalizations requires calmness, control and attention toward the others instead of being self-centered. For France et al. (2001), the non-verbal cues that accompany turn-taking demonstrate mutual attention and responsiveness. According to Bourhis (1982) and Hofstede (1980), some human societies are built upon the development of "speaking well" while others, more communal, favor the knowledge of the social relationships. This is reminiscent of the gradient observed in species of the starling family (Sturnidae). Other communal breeders and group living animals such as the Australian magpies also favor choruses and overlap of songs (e.g., Brown and Farabaugh, 1997). At the other extreme, territorial skylarks have developed continuous songs that prevent turn-taking: the challenger deliberately overlaps the rival and "takes over" (Geberzahn and Aubin, 2014). This recalls some human conversations where the dominant individual disregards the other's turn.

For Takahashi et al. (2013a), vocal turn-taking does not require higher order cognitive capacities. Indeed the temporal features of animal vocal interactions in many ways parallel human communication. In particular, alternating vocal interactions are present in a large number of songbirds while cetaceans and primates seem to have "conversations" (e.g., Snowdon and Cleveland, 1984). However, as mentioned by Snowdon (1982), "in no way do they approach the complexity of human rules…they do indicate that rule-governed communication systems are not unique to humans.

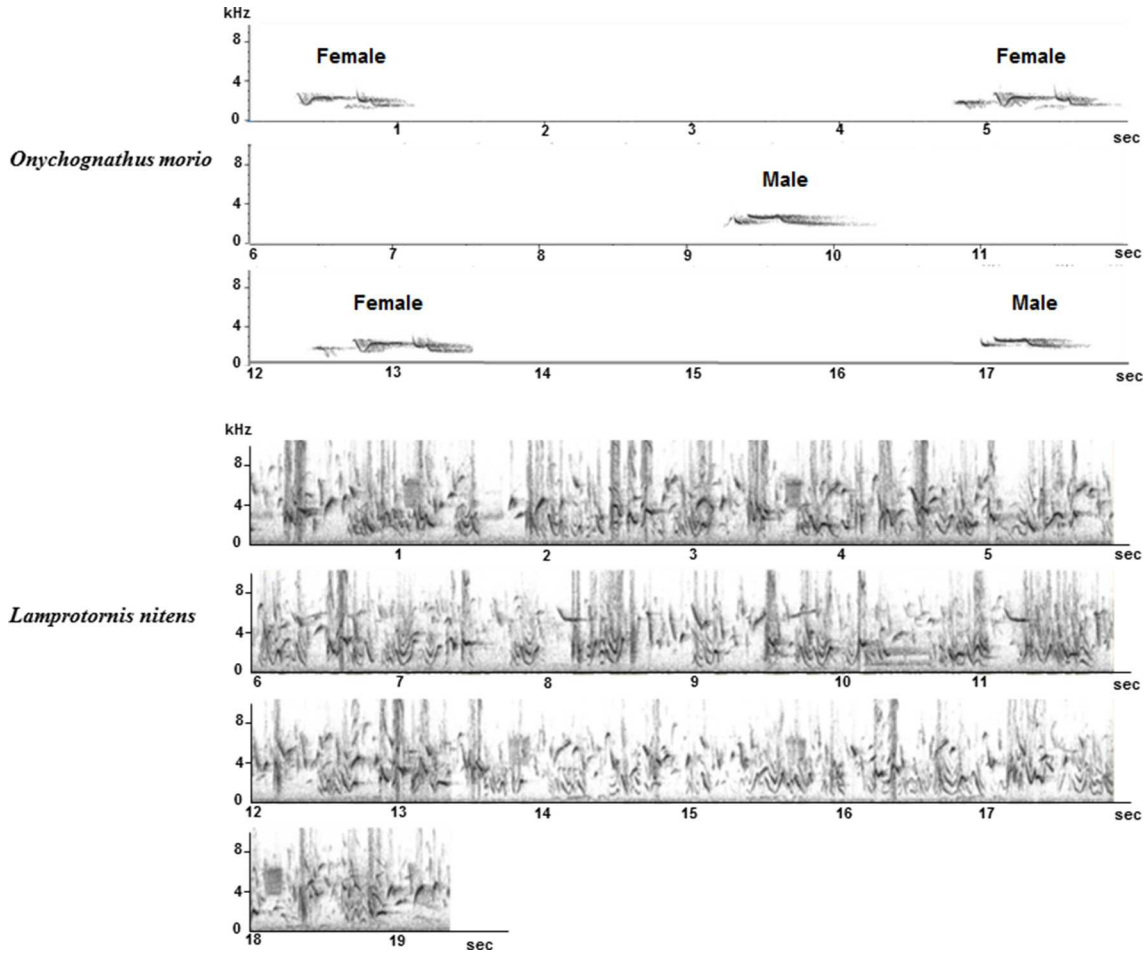


FIGURE 12 | Whistles of a male and a female *O. morio* (Top): whistles are separated by silent intervals. Chorus of *L. nitens*: several birds are singing together and songs overlap.

The use of rule systems for vocal communication is not limited to human beings."

This review makes two additional points: turn-taking is one characteristic feature of human conversations but choruses might well be of interest if the social evolution of language and the intercultural aspects are to be considered; more integrative studies such as those described here (and in progress) for starlings are needed in order to tackle the question of the evolution of rule-governed communication in language.

Acknowledgments

We thank Marion Coulon, Sandrine Alcaix, Hugo Cousillas, Isabelle George, Cécilia Houdelier, Florence Rousseau, Françoise Joubaud, Jean-Pierre Richard, Christophe Lunel, Pat Hulley, Michael Ford, Arjen Van Zweiten and Tshifhiwa Mandiwana-Neudani for their help. This research was funded by programs: OHLL (CNRS), Cognitique (Ministère de la recherche), Studies on African starlings were funded primarily by grants from the CNRS (France) and NRF (South Africa) (CNRS PICS and LIA programs), with additional funding from Rhodes University and Rennes 1 University.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01416

References

Adret-Hausberger, M. (1982). Social influences on the whistled songs of starlings. *Behav. Ecol. Sociobiol.* 11, 241–246. doi: 10.1007/BF00299300

Adret-Hausberger, M. (1986). Temporal dynamics of dialects in the whistled songs of starlings. *Ethology* 71, 140–152. doi: 10.1111/j.1439-0310.1986.tb00579.x

Adret-Hausberger, M. (1989). Song ontogenesis in starlings *Sturnus vulgaris*: are song and subsong continuous? *Bird Behav.* 8, 8–13. doi: 10.3727/015613888791871278

Adret-Hausberger, M., Güttinger, H. R., and Merkel, F. W. (1990). Individual life history and song repertoire changes in a colony of starlings (*Sturnus vulgaris*). *Ethology* 84, 265–280. doi: 10.1111/j.1439-0310.1990.tb00802.x

Adret-Hausberger, M., and Jenkins, P. F. (1988). Complex organization of the warbling song in the European starling *Sturnus vulgaris*. *Behaviour* 107, 138–156. doi: 10.1163/156853988X00322

Albert, E. (1964). Rhetoric, logic and poetics in Burundi: culture patterning of speech and behavior. *Am. Anthropol.* 66, 33–54. doi: 10.1525/aa.1964.66.suppl_3.02a00020

Au, W. W. L., Mobley, J., Burgess, W. C., and Lammers, M. O. (2000). Seasonal and diurnal trends of chorusing humpback whales wintering in waters off western Maui. *Mar. Mammal. Sci.* 16, 530–544. doi: 10.1111/j.1748-7692.2000.tb00949.x

Baker, T. M., Wilson, D. R., and Mennill, D. J. (2012). Vocal signals predict attack during aggressive interactions in black-capped chickadees. *Anim. Behav.* 84, 965–974. doi: 10.1016/j.anbehav.2012.07.022

Bertin, A., Hausberger, M., Henry, L., and Richard-Yris, M.-A. (2007). Adult and peer influences on starling song development. *Dev. Psychobiol.* 49, 362–374. doi: 10.1002/dev.20223

Bertram, B. (1970). The vocal behaviour of the Indian Hill Mynah *Gracula religiosa*. *Anim. Behav. Monogr.* 3, 79–192. doi: 10.1016/S0066-1856(70)80003-6

Biben, M., Symme, D., and Masataka, N. (1986). Temporal and structural analysis of affiliative vocal exchanges in squirrel monkeys. *Behaviour* 98, 259–273. doi: 10.1163/156853986X00991

Bigot, E., Hausberger, M., and Black, J. M. (1995). Exuberant youth: the example of triumph ceremonies in Barnacle geese (*Branta leucopsis*). *Ethol. Ecol. Evol.* 7, 79–85. doi: 10.1080/08927014.1995.9522971

Black, B., and Logan, A. (1995). Links between communication patterns in mother-child, father-child, and child-peer interactions and children's social status. *Child Dev.* 66, 255–274. doi: 10.2307/1131204

Bourhis, R. Y. (1982). "Language policies and language attitudes: le monde de la francophonie," in *Attitudes Towards Language Variation*, eds E. B. Ryan, and H. Giles (London: Edward Arnold), 34–62.

Briefer, E., Rybak, F., and Aubin, T. (2008). When to be a dear enemy: flexible acoustic relationships of neighbouring skylarks, *Alauda arvensis*. *Anim. Behav.* 6, 1319–1325. doi: 10.1016/j.anbehav.2008.06.017

Brockway, B. F. (1969). "Roles of budgerigar vocalization in the integration of breeding behaviour," in *Birds Vocalization*s, ed R. A. Hinde (Cambridge: Cambridge University Press), 131–158.

Brown, E. D., and Farabaugh, S. M. (1991). Song sharing in a group-living songbird, the Australian magpie, *Gymnorhina tibicen*. 3. Sex specificity and individual specificity of vocal parts in communal chorus and duet songs. *Behaviour* 118, 244–274. doi: 10.1163/156853991X00319

Brown, E. D., and Farabaugh, S. M. (1997). "What birds with complex social relationships can tell us about vocal learning: vocal sharing in avian groups," in *Social Influences on Vocal Development*, eds C. T. Snowdon and M. Hausberger (Cambridge: Cambridge University Press), 98–127.

Calame-Griaule, G. (1965). *Ethnologie et Langage. La parole chez les Dogon*. Paris: Galimard.

Camacho-Schlenker, S., Courvoisier, H., and Aubin, T. (2011). Song sharing and singing strategies in the winter wren *Troglodytes troglodytes*. *Behav. Process.* 87, 260–267. doi: 10.1016/j.beproc.2011.05.003

Candiotti, A., Zuberbühler, K., and Lemasson, A. (2012). Convergence and divergence in Diana monkey vocalisations. *Biol. Lett.* 8, 382–385. doi: 10.1098/rsbl.2011.1182

Carter, G. G., Skowronski, M. D., Faure, P. A., and Fenton, B. (2008). Antiphonal calling allows individual discrimination in white-winged vampire bats. *Anim. Behav.* 76, 1343–1355. doi: 10.1016/j.anbehav.2008.04.023

Catchpole, C. K. (1980). Sexual selection and the evolution of complex song among warblers of the genus *Acrocephalus*. *Behaviour* 74, 149–166. doi: 10.1163/156853980X00366

Catchpole, C. K. (1983). Variation in the song of the great reed warbler *Acrocephalus arundinaceus* in relation to mate attraction and territorial defense. *Anim. Behav.* 31, 1217–1225. doi: 10.1016/S0003-3472(83)80028-1

Catchpole, C. K., and Slater, P. J. B. (1995). *Bird Song: Biological Themes and Variations*. Cambridge, England: Cambridge University Press.

Chaiken, M., Boehner, J., and Marler, P. (1993). Song acquisition in European starlings, *Sturnus vulgaris*: a comparison of the songs of live tutored, tape tutored, untutored, and wild caught males. *Anim. Behav.* 46, 1079–1090. doi: 10.1006/anbe.1993.1298

Chen, H. C, Kaplan, G., and Rogers, L. J. (2009). Contact calls of common marmosets (*Callithrix jacchus*): influence of age of caller on antiphonal calling and other vocal responses. *Am. J. Primatol.* 71, 165–170. doi: 10.1002/ajp.20636

Chew, S. J., Vicario, D. S., and Nottebohm, F. (1996). Quantal duration of auditory memories. *Science* 274, 1909–1914. doi: 10.1126/science.274.5294.1909

Chow, C. P., Mitchell, J. F., and Miller, C. T. (2015). Vocal turn taking in a non-human primates is learned during ontogeny. *Proc. R. Soc. B.* 282:20150069. doi: 10.1098/rspb.2015.0069

Clergeau, P. (1989). Estimation des effectifs d'étourneaux reproducteurs et hivernants en France. *L'Oiseau Rev. Fr. Ornithol.* 59, 101–115.

Counsilman, J. J. (1974). Waking and roosting behaviour of the Indian myna. *Emu* 74, 135–148. doi: 10.1071/MU974135

Cousillas, H., George, I., Henry, L., Richard, J. P., and Hausberger, M. (2008). Linking social and vocal brains: could social segregation prevent a proper development of a central auditory area in a female songbird? *PLoS ONE* 3:e2194. doi: 10.1371/journal.pone.0002194

Cousillas, H., Mathelier, M., Richard, J. P., Henry, L., and Hausberger, M. (2006). Social experience: a major factor in the development of a central auditory area. *Naturwissenchaften* 93, 588–596. doi: 10.1007/s00114-006-0148-4

Cousillas, H., Richard, J. P., Mathelier, M., Henry, L., George., I., and Hausberger, M. (2004). Experience-dependent neuronal specialization and functional organization in the central auditory area of a songbird. *Eur. J. Neurosci.* 19, 3343–3352. doi: 10.1111/j.0953-816X.2004.03376.x

Craig, A. (1983). Co-operative breeding in two African starlings, Sturnidae. *Ibis* 125, 114-115. doi: 10.1111/j.1474-919x.1983.tb03089.x

Craig, A. J. F. K. (1987). Co-operative breeding in the Pied Starling. *Ostrich* 58, 176-180. doi: 10.1080/00306525.1987.9633905

Craig, A. J. F. K., and Feare, C. J. (2009). "Family Sturnidae (Starlings)," in *Handbook of the Birds of the World*, Vol. 14, eds J. del Hoyo, A. Elliott, and D. Christie (Barcelona: Lynx Edicions), 654–758.

Craig, A. J. F. K., Hulley, P. E., and Walter, G. H. (1991). The behaviour of Palewinged Starlings and a comparison with other *Onychognathus* species. *Ostrich* 62, 97-108. doi: 10.1080/00306525.1991.9639648

Craig, H. K., and Washington, J. A. (1986). Children's turn-taking behaviours socio-linguistic interactions. *J. Pragmatics* 10, 173–197. doi: 10.1016/0378-2166(86)90086-X

Dabelsteen, T., McGregor, P. K., Holland, J., Tobias, J. A., and Pedersen, S. B. (1997). The signal function of overlapping singing in male robins. *Anim. Behav.* 53, 249–256. doi: 10.1006/anbe.1996.0369

Dahlin, C. R., and Benedict, L. (2013). Angry birds need not apply: a perspective on the flexible form and multifunctionality of avian vocal duets. *Ethology* 119, 1–10. doi: 10.1111/eth.12182

De Groof, G, Poirier, C, George, I, Hausberger, M., and Van der Linden, A. (2013). Functional changes between seasons in the male songbird auditory forebrain. *Front. Behav. Neurosci.* 7:196. doi: 10.3389/fnbeh.2013.00196

Deleporte, P. (2002). Phylogenetics and the aptationist program. Commentary/Andrews et al.: Adaptationism—how to carry out an exaptationist program. *Behav. Brain Sci.* 25, 514–515. doi: 10.1017/S0140525X02310094

Duncan, S. H. (1972). Some signals and rules for taking speaking turns in conversations. *J. Personal. Social Psychol.* 23, 283–292. doi: 10.1037/h0033031

Eens, M. (1997). Understanding the complex song of the European starling: an integrated ethological approach. *Adv. Study Behav.* 26, 355–434. doi: 10.1016/S0065-3454(08)60384-8

Eens, M., Pinxten, R., and Verheyen, R. F. (1989). Temporal and sequential organization of song bouts in European starling. *Ardea* 77, 75–86.

Falls, J. B. (1982). "Individual recognition by sounds in birds," in *Acoustic Communication in Birds*, eds D. E. Kroodsma and E. H. Miller (New York, NY: Academic Press), 237–278.

Falls, J. B., and Brooks, R. J. (1975). Individual recognition of song in the white-throated sparrow. II. *Can. J. Zool.* 53, 1412–1420. doi: 10.1139/z75-170

Feare, C. (1984). *The Starling*. Oxford, UK: Oxford University Press.

Feare, C. J., and Craig, A. (1999). *Starlings and Mynas*. Princeton, NJ: Princeton University Press.

Fedurek, P., Schel, A. M., and Slocombe, K. E. (2013). The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behav. Ecol. Sociobiol.* 67, 1781–1789. doi: 10.1007/s00265-013-1585-7

Feekes, F. (1982). Song mimesis within colonies of *Cacicus c. cela (Leteridae, Aves)*. A colonial password? *Z. Tierpsychol.* 52, 119–152.

Ficken, M. S., Ficken, R. W., and Witkin, S. R. (1978). Vocal repertoire of the blackcapped chickadee. *Auk* 95, 34–48. doi: 10.2307/4085493

Foote, J. R, Fitzsimmons, L. P., Daniel, J., Mennill, D. J., and Ratcliffe, L. M. (2008). Male chickadees match neighbors interactively at dawn: support for the social dynamics hypothesis. *Behav. Ecol.* 19, 1192–1199. doi: 10.1093/beheco/arn087

France, E. F., Anderson, A. H., and Gardner, M. (2001). The impact of status and audio conferencing technology on business meetings. *Int. J. Hum. Comput. Stud.* 54, 857–876. doi: 10.1006/ijhc.2001.0464

Fry, H., Keith, S., and Urban, E. K. (2000). *The Birds of Africa*, Vol. VI. London: Academic Press.

Geberzahn, N., and Aubin, T. (2014). How a songbird with a continuous singing style modulates its song when territorially challenged. *Behav. Ecol. Sociobiol.* 68, 1–12. doi: 10.1007/s00265-013-1616-4

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* 77, 57–76. doi: 10.1017/S1464793101005826

George, I., Alcaix, S., Henry, L., Richard, J. P., Cousillas, H., and Hausberger, M. (2010). Neural correlates of experience-induced deficits in learned vocal communication. *PLoS ONE* 5:e14347. doi: 10.1371/journal.pone.0014347

George, I., Cousillas, H., Richard, J. P., and Hausberger, M. (2003). A new extensive approach to single unit responses using multisite recording electrodes: application to the songbird brain. *J. Neurosci. Methods.* 125, 65–71. doi: 10.1016/S0165-0270(03)00025-6

George, I., Cousillas, H., Richard, J.-P., and Hausberger, M. (2008). A potential neural substrate for processing functional classes of complex acoustic signals. *PLoS ONE* 3:e2203. doi: 10.1371/journal.pone.0002203

George, I., Vernier, B., Richard, J. P., Hausberger, M., and Cousillas, H. (2004). Hemispheric specialization in the primary auditory area of awake and anesthetized starlings (*Sturnus vulgaris*). *Behav. Neurosci.* 118, 597–610. doi: 10.1037/0735-7044.118.3.597

Gérard, J. (1987). "La conversation et les tours de parole, " in *Savoir Parler, Savoir Dire, Savoir Communiquer*, ed J. Gérad (Neuchâtel-Paris: Delacheaux and Niestlé), 33–57.

Gochfeld, M. (1978). Intraspecific social stimulation and temporal displacement of songs in the lesser skylark. *Z. Tierpsychol.* 48, 337–344. doi: 10.1111/j.1439-0310.1978.tb00264.x

Groschupf, K. (1985). Changes in five-striped sparrow song in intra and intersexual conexts. *Wilson Bull.* 97, 102–105.

Hammerschmidt, K., Ansorge, V., Fischer, J., and Todt, D. (1994). Dusk calling in barbary macaques (*Macaca sylvanus*): demand for social shelter. *Am. J. Primatol.* 32, 277–289. doi: 10.1002/ajp.1350320405

Hausberger, M. (1991). The organization of whistle sequences in starlings. *Bird Behav.* 9, 81–88. doi: 10.3727/015613890791749127

Hausberger, M. (1997). "Social influences on song acquisition and sharing in the European starling (*Sturnus vulgaris*)," in *Social Influences on Vocal Development*, eds C. T. Snowdon and M. Hausberger (Cambridge: Cambridge University Press), 128–156.

Hausberger, M., and Black, J. M. (1990). Do females turn males on and off in barnacle goose social display? *Ethology* 54, 232–238. doi: 10.1111/j.1439-0310.1990.tb00799.x

Hausberger, M., and Guyomarc'h, J. C. (1981). Contribution à l'étude des vocalisations territoriales sifflées chez l'étourneau sansonnet *Sturnus vulgaris* en Bretagne. *Biol. Behav.* 6, 79–98.

Hausberger, M., Henry, L., Testé, B., and Barbu, S. (2008). "Contextual sensitivity and birdsong: a basis for social life," in *Evolution of Communicative Flexibility: Complexity, Creativity and Adaptability in Human and Animal Communication*, eds K. Oller and U. Griebel (Cambridge, MA: MIT Press), 121–138.

Hausberger, M., Richard-Yris, M. A., Henry, L., Lepage, L., and Schmidt, I. (1995). Song sharing reflects the social organization in a captive group of European starlings (*Sturnus vulgaris*). *J. Comp. Psychol.* 109, 222–241. doi: 10.1037/0735-7036.109.3.222

Hauser, M. D. (1992). "A mechanism guiding conversational turn-taking in vervet monkeys and rhesus macaques," in *Topics of Primatology*, Vol. 1. *Human Origins*, eds T. Nishida, F. B. M. de Waal, W. McGrew, P. Marler, and M. Pickford (Tokyo: Tokyo University Press), 235–248.

Henry, L., Hausberger, M., and Jenkins, P. F. (1994). The use of song repertoire changes with pairing status in male European starlings. *Bioacoustics* 5, 261–266. doi: 10.1080/09524622.1994.9753256

Hofstede, G. (1980). *Culture Consequences*. Beverly Hills, CA: Sage.

Hooker, T., and Hooker, B. I. (1969). "Duetting," in *Bird Vocalizations*, ed R. A. Hinde (Cambridge: Cambridge University Press), 185–205.

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science* 289, 1355–1357. doi: 10.1126/science.289.5483.1355

Kerbrat-Orecchioni, C. (2001). "Les cultures de la conversation," in *Le Langage, Nature, Histoire et Usage*, ed J. F. Dortier (Auxerre: Editions Sciences Humaines), 209–215.

Koda, H., Lemasson, A., Oyakawa, C., Rizaldi, Pamungkas, J., and Masataka, N. (2013). Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *PLoS ONE* 8:e71432. doi: 10.1371/journal.pone.0071432

Kremers, D., Briseño-Jaramillo, M., Böye, M., Lemasson, A., and Hausberger, M. (2014). Nocturnal vocal activity in captive bottlenose dolphins (*Tursiops truncatus*): could dolphins have presleep choruses? *Anim. Behav. Cogn.* 1, 464–469. doi: 10.12966/abc.11.04.2014

Kroodsma, D. E. (1977). Correlates of song organisation among North American wrens. *Am. Nat.* 111, 995–1008. doi: 10.1086/283228

Kunz, T. H. (1982). "Roosting ecology of bats," in *Ecology of Bats*, ed T. H. Kunz (New York, NY: Plenum Publishing Corporation), 1–55.

Lemasson, A., Gandon, E., and Hausberger, M. (2010). Attention to elders' voice in nonhuman primates. *Biol. Lett.* 6, 325–328. doi: 10.1098/rsbl.2009.0875

Lemasson, A., Glas, L., Barbu, S., Lacroix, A., Guilloux, M., Remeuf, K., et al. (2011). Youngsters do not pay attention to conversational rules: also in nonhuman primates? *Sci. Rep.* 1:22. doi: 10.1038/srep00022

Lemasson, A., Guilloux, M., Rizaldi, Barbu, S., Lacroix, A., and Koda, H. (2013). Age- and sex-dependent contact call usage in Japanese macaques. *Primates* 54, 283–291. doi: 10.1007/s10329-013-0347-5

Locke, J. L. (1993). *The Child's Path to Spoken Language*. Cambridge, MA: Harvard University Press.

Logue, D., and Stivers, T. (2012). Squawk in interaction: a primer of conversation analysis for students of animal communication. *Behaviour* 149, 1283–1298. doi: 10.1163/1568539X-00003031

MacComb, K., and Semple, S. (2005). Co-evolution of vocal communication and sociality in primates. *Biol. Lett.* 1, 381–385. doi: 10.1098/rsbl.2005.0366

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., and Munhall, K. G. (2012).Children's development of self-regulation in speech production. *Curr. Biol.* 22, 113–117. doi: 10.1016/j.cub.2011.11.052

MacWhinney, B. (2008). "Cognitive precursor to language," in *Evolution of Communicative Flexibility: Complexity, Creativity and Adaptability in Human and Animal Communication*, eds K. Oller and U. Griebel (Cambridge, MA: MIT Press), 193–213.

Maltz, D. N., and Borker, R. A. (1982). "A cultural approach to male-female miscommunication," in *Language and Social Identity*, ed J. J. Gumperz (New York, NY: Cambridge University Press), 196–216.

Marler, P. (1970). Bird song and speech development: could there be parallels? *Am. Sci.* 58, 669–673.

Masataka, N., and Biben, M. (1987). Temporal rules regulating affiliative vocal exchanges of squirrel monkeys. *Behaviour* 101, 311–319. doi: 10.1163/156853987X00035

Mennill, D. J., and Ratcliffe, L. M. (2004). Overlapping and matching in the song contexts of black-capped chickadees. *Anim. Behav.* 67, 441–450. doi: 10.1016/j.anbehav.2003.04.010

Miller, P. J. O., Shapiro, A. D., Tyack, P. L., and Solow, A. R. (2004). Call types matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*. *Anim. Behav.* 67, 1099–1107. doi: 10.1016/j.anbehav.2003.06.017

Miura, I (1993). Switching pauses in adult-adult and child-child turn-takings: an initial study. *J. Psycholinguist. Res.* 22, 383–395. doi: 10.1007/BF01068018

Moerman, M. (1977). The preference for self-correction in Thai conversational corpus. *Language* 53, 872–882. doi: 10.2307/412915

Naguib, M. (1990). Effects of song overlapping and alternating in nocturnally singing nightingales. *Anim. Behav.* 88, 1061–1067.

Naguib, M., and Kipper, S. (2005). Effects of different levels of song overlapping on singing behaviour in male territorial nightingales (Luscinia

megarhynchos). *Behav. Ecol. Sociobiol.* 59, 419–426. doi: 10.1007/s00265-005-0066-z

Naguib, M., and Mennill, D. J. (2010). The signal value of birdsong: empirical evidence suggests song overlapping is a signal. *Anim. Behav.* 80, 11–15. doi: 10.1016/j.anbehav.2010.06.001

Naguib, M., Mundry, R., Hultsch, H., and Todt, D. (2002). Responses to playback of whistle songs and normal songs in male nightingales: effects of song category, whistle pitch, and distance. *Behav. Ecol. Sociobiol.* 52, 216–222. doi: 10.1007/s00265-002-0511-1

Poirier, C., Henry, L., Mathelier, M., Lumineau, S., Cousillas, H., and Hausberger, M. (2004). Direct social contacts override auditory information in the song learning process in starlings. *J. Comp. Psychol.* 118, 179–193. doi: 10.1037/0735-7036.118.2.179

Richard, J. P. (1991). Sound analysis and synthesis using an Amiga microcomputer. *Bioacoustics* 3, 45–60. doi: 10.1080/09524622.1991.9753156

Riebel, K. (2003). The "mute" sex revisited: vocal production and perception learning in female songbirds. *Adv. Study Behav.* 33, 49–86. doi: 10.1016/S0065-3454(03)33002-5

Rossano, F. (2013). Sequence organization and timing of bonobo mother-infant interactions. *Interact. Stud.* 14, 160–189. doi: 10.1075/is.14.2.02ros

Rossano, F., and Liebal, K. (2014). "Requests" and "offers" in orangutans and human infants," in *Requesting in Social Interaction,* eds P. Drew and E. Couper-Kuhlen (Amsterdam: John Benjamins), 333–362.

Rowan, M. K. (1955).The breeding biology and behaviour of the redwinged starling *Onychognathus morio*. *Ibis* 97, 663–705. doi: 10.1111/j.1474-919X.1955.tb01926.x

Rutter, D. R., and Durkin, K. (1987). Turn-taking in mother-infant interaction: an examination of vocalization and gaze. *Dev. Psychol.* 23, 54–61. doi: 10.1037/0012-1649.23.1.54

Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematic for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Sankoff, G. (1980). "Multilingualism in Papua New Guinea," in *Social Life of Language*, ed G. Sankoff (Philadelphia: University of Pennsylvania Press), 95–132.

Schegloff, E. A. (2000). Overlapping table and the organization of turn-taking for conversation. *Lang. Soc.* 29, 1–63. doi: 10.1017/S0047404500001019

Schmidt, R., Kunc, H. P., Amrhein, V., and Naguib, M. (2006). Responses to interactive playback predict future pairing success in nightingales. *Anim. Behav.* 72, 1355–1362. doi: 10.1016/j.anbehav.2006.03.022

Schulz, T. M., Whitehead, H., Gero, S., and Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. *Anim. Behav.* 76, 1977–1988. doi: 10.1016/j.anbehav.2008.07.032

Searcy, W. A., and Beecher, M. D. (2009). Song as an aggressive signal in songbirds. *Anim. Behav.* 78, 1281–1292. doi: 10.1016/j.anbehav.2009.08.011

Searcy, W. A., and Beecher, M. D. (2011). Continued scepticism that song overlapping is a signal. *Anim. Behav.* 81, e1–e4. doi: 10.1016/j.anbehav.2010.10.013

Sidnell, J. (2001). Conversational turn-taking in a Caribbean English Creole. *J. Pragmatics* 33, 1263–1290. doi: 10.1016/S0378-2166(00)00062-X

Snowdon, C. T. (1982). "Linguistic and psycholinguistic approaches to primate communication," in *Primate Communication*, eds C. T. Snowdon, C. H. Brown, and M. R. Pedersen, (Cambridge, England: Cambridge University Press), 212–238.

Snowdon, C. T., and Cleveland, J. (1984). "Conversations" among pygmy marmosets. *Am. J. Primatol.* 7, 15–20. doi: 10.1002/ajp.1350070104

Soltis, J., Leong, K., and Savage, A. (2005). African elephant vocal communication I: antiphonal calling behaviour among affiliated females. *Anim. Behav.* 70, 579–587. doi: 10.1016/j.anbehav.2004.11.015

Sontag, W. A. (1991). Habitatsunterschiede, Balzverhaten, Paarbildung und Parbindung beim Lappenstar *Creatophora cinerea*. *Acta Biol. Benrodi* 3, 99–114.

Sorjonen, J. (1983). Transmission of the two most characteristic phrases of the song of the thrush nightingale *Luscina luscina* in different environmental conditions. *Ornis Scand.* 14, 278–288. doi: 10.2307/3676320

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Sugiura, H. (2007). Adjustment of temporal call usage during vocal exchange of coo calls in Japanese macaques. *Ethology* 113, 528–533. doi: 10.1111/j.1439-0310.2007.01349.x

Symmes, D., and Biben, M. (1988). "Conversational vocal exchanges un squirrel monkeys," in *Primate Vocal Communication*, eds D. Todt, P. Goedeking, and D. Symmes (Berlin: Springer), 123–132.

Takahashi, D. Y., Narayanan, D., and Ghazanfar, A. A. (2013b). "Development of self-monitoring essential for vocal interactions in marmoset monkeys," in *IEEE Conference on Development and Learning and Epigenetic Robotics* (Osaka).

Takahashi, D. Y., Narayanan, D. Z., and Ghazanfar, A. A. (2013a). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* 23, 2162–2168. doi: 10.1016/j.cub.2013.09.005

Thieltges, H., Henry, L., Biquand, V., and Deleporte, P. (2014). Short-term variations of dialects in short songs of two species of colonial caciques (*Cacicus*). *Acta Acustica United Acustica* 100, 759–766. doi: 10.3813/AAA.918755

Thorpe, W. H. (1961). *Bird Song*. Cambridge, UK: Cambridge University Press.

Tinbergen, N. (1963). On aims and methods of ethology. *Z. Tierpsychol.* 20, 410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x

Todt, D., and Hultsch, H. (1982). Impairment of vocal signal exchange in the monogamous duet-singer *Cossypha heuglini* (Turdidae): effects of pairbond maintenance. *Ethology* 60, 265–274.

Todt, D., and Naguib, M. (2000). Vocal interactions in birds: the use of song as a model in communication. *Adv. Study Behav.* 29, 247–296. doi: 10.1016/S0065-3454(08)60107-2

Trainer, J. M., McDonald, D. B., and Learn, W. A. (2002). The development of coordinated singing in cooperatively displaying long-tailed manakins. *Behav. Ecol.* 13, 65–69. doi: 10.1093/beheco/13.1.65

Tyack, P. L. (2000). Dolphins whistle: a signature tune. *Science* 289, 1310–1311. doi: 10.1126/science.289.5483.1310

Verheyen, R. F. (1980). "Breeding strategies of the starling," in *Bird Problems in Agriculture*, eds E. N. Wright, I. R. Inglis, and C. J. Feare (Croydon, UK: British Crop Protection Council), 69–82.

Versace, E., Endress, A. D., and Hauser, M. D. (2008). Pattern recognition mediates flexible timing of vocalizations in nonhuman primates: experiments with cottontop tamarins. *Anim. Behav.* 76, 1885–1892. doi: 10.1016/j.anbehav.2008.08.015

Voigt, C., Leitner, S., and Gahr, M. (2006). Repertoire and structure of duet and solo songs in cooperatively breeding white-browed sparrow weavers. *Behaviour* 143, 159–182. doi: 10.1163/156853906775900739

Wasserman, F. E. (1977). Intraspecific acoustical interference in the White-throated Sparrow (*Zonotrichia albicollis*). *Anim. Behav.* 25, 949–952. doi: 10.1016/0003-3472(77)90045-8

Yamaguchi, C., Izumi, A., and Nakamura, K. (2009). Temporal rules in vocal exchanges of phees and trills in common marmosets (*Callithrix jacchus*). *Am. J. Primatol.* 71, 617–622. doi: 10.1002/ajp.20697

Yosida, S., Kobayasi, K. I., Ikebuchi, M., Ozaki, R., and Okanoya, K. (2007). Antiphonal Vocalization of a Subterranean Rodent, the Naked Mole-Rat (*Heterocephalus glaber*). *Ethology* 113, 703–710. doi: 10.1111/j.1439-0310.2007.01371.x

Yosida, S., and Okanoya, K. (2009). Naked mole−rat is sensitive to social hierarchy encoded in antiphonal vocalization. *Ethology* 115, 823–831. doi: 10.1111/j.1439-0310.2009.01677.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

CORRECTION
published: 16 December 2015
doi: 10.3389/fpsyg.2015.01924

Corrigendum: Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model

Laurence Henry[1]*, *Adrian J. F. K. Craig*[2], *Alban Lemasson*[1,3] and *Martine Hausberger*[1]

[1] Laboratoire d'éthologie animale et humaine, Centre National de la Recherche Scientifique, UMR 6552, Université de Rennes 1, Rennes, France, [2] Department of Zoology and Entomology, Rhodes University, Grahamstown, South Africa, [3] Laboratoire d'éthologie animale et humaine, Centre National de la Recherche Scientifique, UMR 6552, Station Biologique, Université de Rennes 1, Paimpont, France

Keywords: turn-taking, vocal interactions, conversation rules, mammals, birdsong, sturnids

A corrigendum on:

Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model
by Henry, L., Craig, A. J. F. K., Lemasson, A., and Hausberger, M. (2015). Front. Psychol. 6:1416. doi: 10.3389/fpsyg.2015.01416

Figure 3 of the article by Henry et al. (2015) contained a minor error, which we correct here.

OPEN ACCESS

Edited and reviewed by:
Marisa Casillas, Max Planck Institute for Psycholinguistics, Netherlands

***Correspondence:**
Laurence Henry
lhenry@univ-rennes1.fr

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *22 October 2015*
Accepted: *30 November 2015*
Published: *16 December 2015*

Citation:
Henry L, Craig AJFK, Lemasson A and Hausberger M (2015) Corrigendum: Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model. Front. Psychol. 6:1924. doi: 10.3389/fpsyg.2015.01924

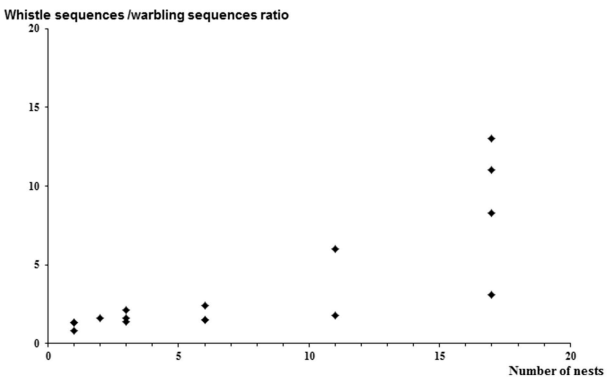


FIGURE 3 | Whistle sequences/warbling sequences ratio according to the number of nests per colony ($rs = 0.9$, $p = 0.003$). Starlings produce more discontinuous song (whistles) when the number of neighbors is high (dense colonies).

Figure captions 4, 5, and 12 contained minor errors, which we correct here.

Figure 4. Song style of birds belonging to colonies of different sizes. Although the birds were recorded in very different conditions, a clear trend appeared toward an increase in whistling (hence discontinuous songs) and a decrease of warbling (hence continuous song) with increasing colony size (= number of neighbors). X: mean number of whistles per sequence (From Hausberger, 1997).

Figure 5. Intervals separating two successive whistles produced by two different individuals during vocal interaction (overlap: when two whistles overlap). Most whistling exchanges show an interval of 2 s or less between the first and second whistle (arrow).

Figure 12. Whistles of a male and a female *O. morio* (Top): whistles are separated by silent intervals. Choruses of *L. nitens*: several birds singing together with their songs in overlap.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 02 February 2015
doi: 10.3389/fpsyg.2015.00089

Anticipation in turn-taking: mechanisms and information sources

Carina Riest*, Annett B. Jorschick and Jan P. de Ruiter

Faculty for Linguistics and Literary Studies, Bielefeld University, Bielefeld, Germany

Edited by:
Kobin H. Kendrick, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Martin John Pickering, University of Edinburgh, UK
Sara Bögels, Max Planck Institute for Psycholinguistics, Netherlands

****Correspondence:***
Carina Riest, Faculty for Linguistics and Literary Studies, Bielefeld University, PF 100131, 33615 Bielefeld, Germany
e-mail: carina.groos@uni-bielefeld.de

During conversations participants alternate smoothly between speaker and hearer roles with only brief pauses and overlaps. There are two competing types of accounts about how conversationalists accomplish this: (a) the signaling approach and (b) the anticipatory ('projection') approach. We wanted to investigate, first, the relative merits of these two accounts, and second, the relative contribution of semantic and syntactic information to the timing of next turn initiation. We performed three button-press experiments using turn fragments taken from natural conversations to address the following questions: (a) Is turn-taking predominantly based on anticipation or on reaction, and (b) what is the relative contribution of semantic and syntactic information to accurate turn-taking. In our first experiment we gradually manipulated the information available for anticipation of the turn end (providing information about the turn end in advance to completely removing linguistic information). The results of our first experiment show that the distribution of the participants' estimation of turn-endings for natural turns is very similar to the distribution for pure anticipation. We conclude that listeners are indeed able to anticipate a turn-end and that this strategy is predominantly used in turn-taking. In Experiment 2 we collected purely reacted responses. We used the distributions from Experiments 1 and 2 together to estimate a new dependent variable called Reaction Anticipation Proportion. We used this variable in our third experiment where we manipulated the presence vs. absence of semantic and syntactic information by low-pass filtering open-class and closed class words in the turn. The results suggest that for turn-end anticipation, both semantic and syntactic information are needed, but that the semantic information is a more important anticipation cue than syntactic information.

Keywords: turn-taking, timing, anticipation, reaction, conversation

INTRODUCTION

Participants in a conversation have a number of tasks that they have to perform simultaneously. They have to comprehend the speaker's utterance while at the same time they need to prepare their response to that utterance, preferably before the current speaker ends their turn. Despite the complexity of these processes the alternation between the speaker and the hearer roles is generally timed with only short pauses and overlaps (Sacks et al., 1974). This conversational phenomenon is an important part of the turn-taking organization.

There are two competing main approaches providing an explanation for the turn-taking organization: the *anticipatory* approach, in which it is assumed that participants are able to predict the end of a turn in advance, and the *signaling* approach, which assumes that listeners perceive specific signals to detect the end of a turn.

The aim of this study was first to determine the relative contribution of these two proposed mechanisms to turn-taking and second, to investigate which linguistic information sources listeners predominantly use for end-of-turn anticipation. To this end, we conducted a series of button-press experiments with turns from natural conversations while manipulating both the respective critical information sources and the task.

The anticipatory approach argues that the precise timing in conversations can only be explained by the listeners' ability to make accurate predictions about the end of the speaker's utterances. Depending on the assumed anticipatory model listeners use various kinds of information to anticipate. The first to claim that listeners are able to anticipate a turn ending were Sacks et al. (1974). In their famous and often-cited turn-taking model they provide an explanation for the characteristic smooth speaker transitions in natural conversation. According to their model, turns consist of syntactic building blocks called turn-constructional units. Listeners are able to predict the end of a turn-constructional unit. At this point a speaker change becomes relevant. This point in time is called a transition-relevance place. When a turn arrives at a transition-relevance place it is possible (a) for the current speaker to select another speaker, or (b) for another speaker to self-select and start talking. If neither option (a) nor (b) is used the current speaker can produce another turn.

In contrast, the signaling approach assumes that turn transitions are regulated by an exchange of conventional vocal or gestural signals (e.g., Yngve, 1970). So in this approach, participants in a conversation do not *anticipate* these signals but *react* to them after having perceived them. Influential proponents of the signaling

approach who did numerous studies on finding explicit turn taking signals are Duncan (1972, 1973), Duncan and Niederehe (1974), and Duncan and Fiske (1977). They assume that there exist definite signals that are displayed and responded to according to specific rules. According to Duncan (1972) such signals are composed of one or more of six behavioral cues: (1) any phrase-final intonation other than sustained, intermediate pitch level, (2) drawl on the final syllable or on the stressed syllable of a terminal clause, (3) the termination of any hand gesticulation, (4) sociocentric sequences (stereotyped expressions like "you know," "isn't it," etc.), (5) drop in pitch and/or loudness in conjunction with one of the sociocentric expressions, or (6) termination of a grammatical clause. According to Duncan and Fiske (1977) speakers always produce at least one of these turn transition cues at the end of their turn, to which listeners react by initiating their next turn. The more cues a speaker produces the more likely a change of speaker role is at that point.

The standard argument against the signaling approach is that the relevant cues occur too late in the speaker's turn to enable timely speaker changes. As a counter-argument, Heldner and Edlund (2010) note that the timing of floor changes is not as precise as it is often claimed. In their analysis of three different conversational corpora 41–45% of between-speaker intervals were longer than 200 ms. They claim that these intervals are potentially long enough for people to react to end-of-turn signals. Their argumentation is based on the distribution of observed delays and pauses in conversational turn-transfers. In their view, pauses longer than 200 ms could also plausibly be explained by assuming they were reactions to signals (p. 566), while pauses shorter than 200 ms could correspond to anticipation (55–59% of the turn transitions in the investigated corpora). Their reaction threshold explanation is based on minimal response times, which were investigated under maximally favorable conditions. Their argument for this strict threshold is that interlocutors are highly trained to recognize gaps, when they can start their turn. But even if one assumes higher thresholds reaching up to 600 ms (Jescheniak et al., 2003; Indefrey and Levelt, 2004; Schnur et al., 2006) Heldner and Edlund (2010) argue that the proportion of responses which can be explained by reaction would be lower, but would not be eliminated.

We want to suggest that the presence of gaps longer than 200 ms does not necessarily mean that the turn before the gap was reacted to. Speakers often intentionally delay the production of so-called 'dispreferred' responses, which leads to longer pauses (see, e.g., Levinson, 1983; Kendrick and Torreira, 2014). So pauses longer than 200 ms are not necessarily caused by reaction, but can also be caused by an anticipated response that was nevertheless intentionally delayed. Conversely, response times of shorter than 200 ms need not always be caused by anticipation, but can be early reactions to perceived signals (false alarms). Hence, using a fixed cut-off value does not give us an accurate estimate of the relative number of anticipated and reacted turn transitions.

One possible criticism regarding the anticipatory approach is that Sacks et al. (1974) do not explain the mechanism responsible for anticipation, and more specifically, which information listeners use to 'project' when a turn is going to end (Sacks et al., 1974; Power and Dal Martello, 1986; O'Connell et al., 1990). Sacks et al. (1974) present only observational evidence suggesting that syntax and intonation play an important role in this process. But in the last decade possible mechanisms of turn-end anticipation have been investigated in more depth.

To investigate the role of intonational contour and lexico-syntactic cues in end-of-turn anticipation De Ruiter et al. (2006) performed a button press experiment presenting turns taken from natural Dutch conversations to participants. The instruction was to press a button when they thought the turn was going to end. They presented unaltered turns as well as manipulated turns where the lexico-syntactic information was absent but the intonational contour remained intact and vice versa. The intonational contour was manipulated by completely flattening the pitch leaving duration, rhythm and intensity intact. The lexico-syntactic information was manipulated by low-pass filtering the original turn fragment. In this way, words could no longer be identified, but the pitch contour remained intact. The results show that for unaltered turns, the average response time was about 200 ms before the turn was finished. This indicates that rather than waiting for the end of the turn and then react, the participants tried to anticipate the turn ending. With intonation contour absent but intact lexico-syntactic information, the participants were still able to accurately anticipate the turn ending. But the anticipation accuracy deteriorated significantly in absence of the lexico-syntactic information. The authors concluded that the lexico-syntactic structure is necessary (and perhaps even sufficient) for accurate end-of-turn projection. They suggested that the syntactic structure provides constraining information about the upcoming words and serves as a temporal resource for the listeners to monitor the unfolding turn. An important difference between the task used by De Ruiter et al. (2006) and turn-taking in natural communication is that listeners do not need to prepare and produce an utterance. This actually led to more accurate responses in the experiment compared to the responses in the natural conversations from which the experimental stimuli were culled. Hence, we believe that the results from this methodology are at least qualitatively generalizable to the natural situation.

Keitel et al. (2013) used eye-tracking methodology to investigate the influence of semantic content and intonation on anticipation ability during development. They presented recordings of actors performing conversations to three different age groups (prelinguistic 6–12 months, linguistic 24–36 months, adults) while measuring their gaze. The conversations were presented either with normal or flattened intonation. If a gaze was shifted from the current to the next speaker at least 500 ms before the end of the current turn, it was considered anticipatory. But if the gaze shifted after the listener began to speak the gaze shift was coded as reactive. The results showed that in contrast to younger infants, children at the age of three are already able to reliably anticipate the end of turns. Furthermore, intonation influenced anticipation only in this specific age group, suggesting that at that age they rely more strongly on intonational information for anticipation than adults. The authors explained this finding by noting that the syntactic and semantic competence of the 3-year-olds is not yet adult-like. This is in line with the finding that adults tend to rely on prosody for the

detection of turn-ends only when neither semantic nor syntactic information is available (Grosjean and Hirt, 1996).

A comparable study was done by Casillas and Frank (2013) who also investigated which linguistic cues children use to anticipate a turn ending. In contrast to Keitel et al. (2013) they tested 1–7 year-olds and instead of using conversations done by actors, they measured the children's gaze shifts while watching videos of conversations between puppets. Casillas and Frank (2013) found that even 1 and 2-year-olds anticipated turn endings, and that their anticipation correlated positively with the duration of the gap between two successive turns. They also manipulated the prosodic or lexical information (or both) of the conversations, and compared question with non-question turns. In their general discussion, they write that "Question effects are strongest when *both* prosodic and lexical cues are present, contrary to prior findings with adult listeners that found lexical information sufficient to predict upcoming turn-end boundaries (De Ruiter et al., 2006)" (emphasis in original). We are not convinced that there is a clear contradiction between their study and the result of De Ruiter et al. (2006) for the following reasons. First, the study by Casillas and Frank (2013) does not provide enough information to assess whether there is a statistically significant effect corresponding to this specific claim. Second, in the study by De Ruiter et al. (2006), the factor Question vs. No-Question was not investigated. (In Stivers et al. (2009) the data from De Ruiter et al. (2006) was reanalyzed and indeed showed no difference between responses to questions and non-questions, but that was only for the natural data.) Finally, it is possible, perhaps even plausible, that asking actors to record a conversation speaking "as if they were on a children's television show" (p. 2) will result in prosodic patterns that are more exaggerated than in natural speech, due to the explicit child-directedness of the actors' speech. For these reasons, we do not (yet) see a clear contradiction between the results of Casillas and Frank (2013) and those of De Ruiter et al. (2006).

To investigate how listeners use lexico-syntactic information to anticipate turn-ends Magyari and De Ruiter (2012) conducted a gating study. They used the experimental stimuli of De Ruiter et al.'s (2006) study and selected turns of which the ends were either predicted with a high or with a low accuracy in the button-press experiment. The results showed that the proportion of the correct guesses of upcoming words was higher when the accuracy of button-press in the original experiment was higher. Furthermore, in the gating study the participants expected more words to come with those turns that resulted in button presses that occurred too late in De Ruiter et al.'s (2006) study. They concluded that listeners make predictions in advance about *which*, and therefore *how many*, words will follow in a turn. These predictions help to estimate the remaining duration of the turn.

The idea that lexico-syntactic information serves as source for listeners' anticipation performance is also supported by conversation-analytic studies (e.g., Ford and Thompson, 1996; Selting, 1996; Caspers, 2003). Caspers (2003) showed in her quantitative investigation that turn transitions are always located at syntactic completion points. She concluded that syntax constitutes the main information source for end-of-turn projection. Similar findings, based on a quantitative analysis of standard German, have been presented in Selting (1996), who concluded that listeners primarily exploit syntactic structure to project turn endings. Ford and Thompson (1996) found in their analysis of an American English face-to-face corpus that speaker change most frequently occurred when syntactic completion was combined with intonational as well as pragmatic completion. They concluded that syntax operates together with intonation and pragmatics to project the end of turns (see also Gravano and Hirschberg, 2011). As not all these studies found a perfect correspondence of syntactic completion points to turn-transitions, it remains an intriguing question how the distinction between those syntactic completions that are, and those that aren't treated as turn-ends by the listeners is made. Unfortunately, this question cannot be satisfactorily answered by studying correlations in dialog corpora, but would require explicit experimentation to be able to distinguish correlation from causation.

To summarize, there is evidence from multiple sources that listeners are able to anticipate the end of the speaker's turn (De Ruiter et al., 2006; Casillas and Frank, 2013; Keitel et al., 2013). But the mere existence of an anticipation ability does not imply that it is actually used to predict when a turn is finished in natural communication. Furthermore, Heldner and Edlund (2010) argued that turn-taking could at least partially be explained by assuming that conversationalists simply react to signals. Thus, the first question we want to investigate in this study is: is turn-taking based on anticipation or on reaction?

EXPERIMENT 1

To determine the relative role of anticipation and reaction in turn-taking we conducted a button-press experiment using the same experimental methodology as in De Ruiter et al. (2006). We took turns from natural conversations and asked the participants to indicate the end of the turn by pressing a button. In the turns, we manipulated the information available for anticipation of the turn end and studied the effect of this manipulation on the projection accuracy. Our manipulations ranged from providing complete advance information about the turn-end to completely removing all linguistic information from the turn. (These manipulations are described in detail below.) The logic is that if the projection accuracy in responding to the original (unchanged) turns is comparable to responses to turns with advance information, then this is evidence for anticipation. On the other hand, if the projection performance to the natural turns is similar to the responses to the turns without or with substantially reduced linguistic information, this is evidence for people reacting to the perceived end of the turn.

MATERIALS AND METHODS

Compliance with ethics guidelines

The experimental methods used in this project have been approved by the Ethics Board of Bielefeld University. Informed consent was obtained from all subjects.

Participants

Eighty native speakers of German participated in Experiment 1 (56 females, 24 males).

Stimulus collection

The stimulus collection procedure is the same as the one described in De Ruiter et al. (2006). For maximum ecological validity we took our stimuli from a natural German 'telephone' corpus (audio-only conversation), which we recorded in our lab. We recorded 16 native speakers of German in eight dyadic conversations (four female–male, three female–female, one male–male). The participants in each dyad were friends. For the stimulus collection we told the participants to just talk about anything they liked and gave them no further instruction. Each dyad's conversation lasted 20 min, resulting in a total of 160 min of recorded conversation.

For the audio recordings we put the participants in two separate rooms and required them to wear closed headphones. Directional microphones were placed on a table in front of them. We established a telephone-like connection between them, such that both participants could hear both themselves and their interlocutor. The speech of each of the two participants was recorded separately on the two channels of a stereo recording device. This way, we avoided cross talk between the participants in our recordings. The participants rapidly got used to the recording situation and the resulting conversations appeared natural and lively.

After recording the corpus, the conversations were transcribed, registering overlaps, pauses, laughter, turn beginnings and endings, assessments (Goodwin, 1986), and continuers (Schegloff, 1982). In addition we measured the Floor Transfer Offset (FTO) of 1597 turn transitions. The FTO value is defined "as the difference (in seconds) between the time that turn starts and the moment the previous turn ends" (De Ruiter et al., 2006, p. 516). Hence, a gap between two turns is characterized by a positive FTO value and an overlap by a negative one. **Figure 1** shows the distribution of the FTO values.

Although the general shape of the FTO distribution resulting from the German telephone corpus looks similar to the Dutch FTO distribution from De Ruiter et al. (2006), the distributions differ in a number of aspects. There are small differences in the means, variances, skewness, and kurtosis (see **Table 1**)[1].

Table 1 | Comparison of Dutch and German telephone corpora.

		Dutch telephone FTO	German telephone FTO
N		1507	1597
Mean	[ms]	0	131
Median	[ms]	38	141
Mode	[ms]	173	162
Variance	[ms]	338	234
Minimum	[ms]	−3080	−2955
Maximum	[ms]	2839	2902
Skewness		−0.348	0.136
Kurtosis		6.923	3.124

From this corpus we randomly selected 100 target turns and an additional 16 turns for practice purposes. We took care that the turns contained at least five words so that the participants in the planned button-press experiments obtained enough information content to potentially base their reaction on. Furthermore, we made sure that the random selection reflected the distribution of pauses and overlaps of the natural conversations. Furthermore we balanced the sex of the speaker in the target turns (50 % female, 50% male). The total number of different speakers in our target stimuli was 16. **Table 2** presents some descriptive statistics of the target turns.

After selecting the target turns, we extracted them into individual sound files using Praat (Boersma and Weenink, 2012) and created four different versions of each stimulus. These versions were as follows.

Natural-Turn. The target turn was presented as it occurred in the natural conversations. In this condition the participants had access to all potentially relevant information to base their anticipation or reaction on.

Advance-Knowledge. The participants could first read the content (a literal transcription) of the turn before they heard the target stimulus. Because the participants knew in advance how the turn

[1]In order to be able to meaningfully compare the higher moments of the two distributions, three outliers containing unrealistically large positive FTO values were removed from the data from the experiment by De Ruiter et al. (2006) for this table.

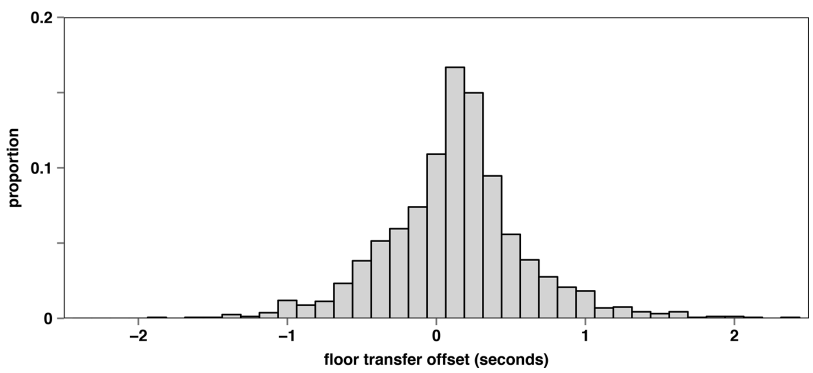


FIGURE 1 | Floor Transfer Offset (FTO) distribution of the German telephone corpus.

Table 2 | Descriptive statistics of target turns.

		Minimum	Maximum	Mean	Mode	SD
Duration	[ms]	863	7105	3157	3136	1415
FTO	[ms]	−1828	1257	96	−70	417
Number of Words		5	29	13	8	6

was going to end, they were, in principle, maximally capable to anticipate the turn end. In this condition the response distribution of anticipated responses was measured.

Scrambled-Word-Order. We randomly changed the order of the words within the target turn using Praat. The pauses between the words in the original were assigned to the subsequent word. The resulting stimuli therefore had the same duration as the Natural-Turn stimuli. In this condition there was no sequential word-order information to base the anticipation on, but there were still words present. Thus, the predictability of a word on the basis of its preceding words is switched off, i.e., the cloze probability (Taylor, 1953) of the words in the resulting turns was very low. In contrast to the Natural-Turn condition the anticipation of the turn end on the basis of sequential lexical information was made impossible.

Noise. The Noise condition was created using a Praat script that convolved the speech stimulus of the natural turn with white noise. The resulting sample of constant noise had the same duration and frequency spectrum as the original fragment. This condition served as a comparative baseline from which all linguistic information that could be used for anticipation was removed. The only way to be certain that the turn has ended in this condition is to react to the turn end. This condition measured the response distributions when the participants had no choice but to react to the end of the turn.

In order to control for subjective loudness between conditions and stimuli we adjusted the loudness of all stimuli to a reference sone value.

Design

Each participant was presented with four trial blocks (Natural turn, Advance-Knowledge, Scrambled-Word-Order, Noise) each containing 25 target turns. Within each block there were four practice trials followed by the 25 target turns. We created eight different experimental lists. In the first four lists we permutated the block order according to a Latin-square design. The remaining four lists were the same as the first four lists with the block order as well as the presentation order of the stimuli reversed. Each of the target turns appeared in all four conditions across the lists but none appeared twice within the same experimental list.

Procedure

The participants received a written instruction that they had to listen to short audio fragments, taken from real conversations, and to press a button as soon as they thought the speaker in the fragments would finish speaking. They were informed that they would be presented with four different blocks, and that in one of these blocks they had to first read the content of the fragment before they heard the corresponding audio fragment. Furthermore, they were informed that in two blocks the stimuli were manipulated acoustically. The stimuli were presented to them via closed headphones. We randomly assigned the participants to one of the eight experimental lists (10 per list).

The participants were presented first with the four practice trials and after that with the corresponding trial block. After each practice block the participants got the chance to ask the experimenter questions. Each experimental block contained a visual countdown from 3 to 1 followed by the auditory presentation of the stimuli. As soon as the participants pressed the button the sound was immediately cut off. In this way we made sure that the participants got no feedback about their performance. The trial block Advance-Knowledge differed from the other trial blocks because after the visual countdown the participants were presented with a written sentence, representing the content of the turn. After pressing the button the sentence disappeared and the acoustic presentation of the stimulus started.

For the presentation of the stimuli we used the E-Prime software package (Schneider et al., 2012a,b), which also allowed us to record the time from stimulus onset to button press.

Results and discussion

We first calculated the BIAS, which is defined as response time minus the duration of the target turn. **Figure 2** shows the BIAS distributions for the four different conditions. **Figure 3** shows an overview of the average BIAS per condition. The average BIAS is negative in all conditions, which gives a first hint that participants tried to anticipate the turn ending, rather than wait until the turn fragment was over.

An ANOVA for the dependent variable BIAS showed a significant effect for presentation condition (by subjects: $F1(3,315) = 23.259$, $p < 0.001$; by items: $F2(3,297) = 18.82$, $p < 0.001$). Bonferroni-corrected paired t-tests, pairing over identical turn fragments from the two conditions under comparison, revealed that the Natural turn condition led to significantly more negative BIAS than the Noise and the Scrambled-Word-Order condition. The latter condition led to significantly more negative BIAS than the Noise condition. Whereas the BIAS in the Advance-Knowledge and the Natural turn condition did not differ significantly from each other.

Conventional significance tests are designed to reject the null hypothesis without fault in the limit of infinite sample size. This is characterized by vanishing p-values and unbounded t-values. In contrast, if the null hypothesis is true and infinite sample sizes are considered the p-values are not converging to any limit value. Correspondingly, under the null hypothesis, all p-values are all equally likely (Rouder et al., 2009). Hence, it is not possible to claim evidence favoring a null hypothesis using conventional significance tests. We therefore also performed a Bayesian analysis (Jeffreys, 1961; Kass and Raftery, 1995) for the Advance-Knowledge and the Natural-Turn condition by comparing them using a Bayesian paired t-test (Rouder et al., 2009). To be consistent with Morey and Rouder (2011) and Rouder et al. (2012) we used a Cauchy

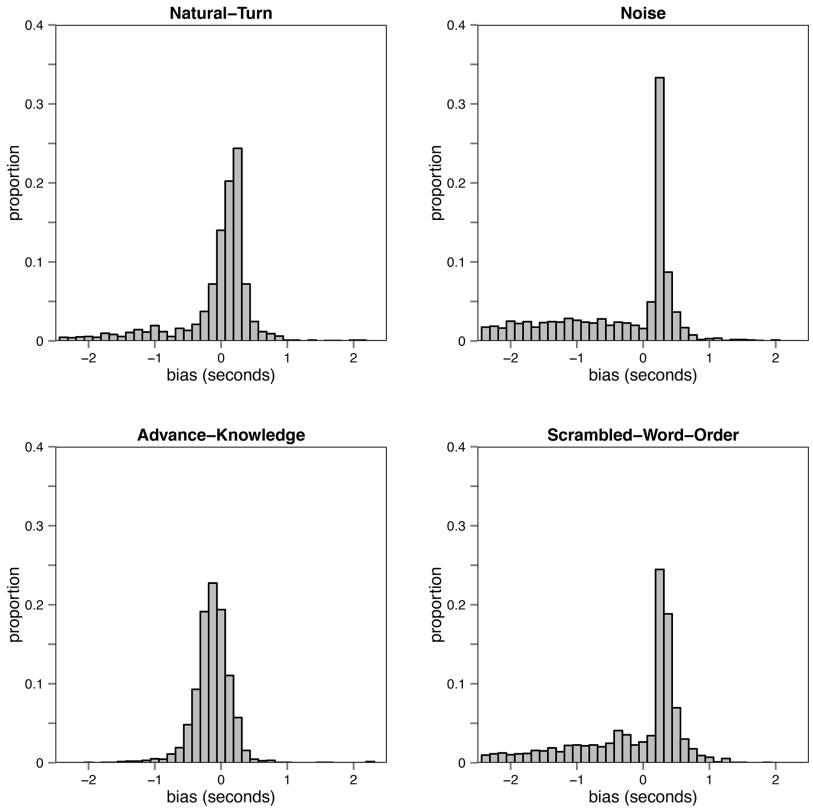


FIGURE 2 | Response distributions per condition from Experiment 1.

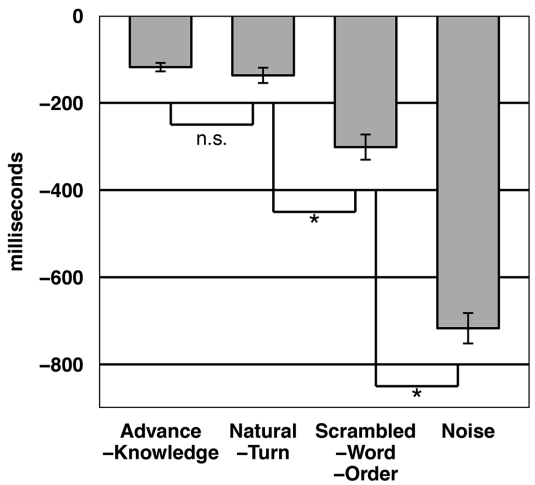


FIGURE 3 | Average BIAS of responses per condition as measured in Experiment 1. Asterisk indicates statistical significance at the 0.05 level.

prior with scale parameter $\sqrt{2}$ for the standardized effect size in combination with a Jeffreys prior on the variance. The analysis was performed using the BayesFactor package (Morey et al., 2014) for R (R Development Core Team, 2009). An overview of a common textual interpretation of Bayes Factor values is presented in **Table 3**.

Table 3 | Evidence Categories for Bayes Factor, adapted from Jeffreys (1961), cited in Wetzels et al. (2011).

Bayes factor	Interpretation
>100	Decisive evidence for H_A
30–100	Very strong evidence for H_A
10–30	Strong evidence for H_A
3–10	Substantial evidence for H_A
1–3	Anecdotal evidence for H_A
1	No evidence
1/3–1	Anecdotal evidence for H_0
1/10–1/3	Substantial evidence for H_0
1/30–1/10	Strong evidence for H_0
1/100–1/30	Very strong evidence for H_0
<1/100	Decisive evidence for H_0

The Bayesian paired *t*-test using item means for the variable BIAS revealed that the null hypothesis, stating that Advance-Knowledge and Natural-Turn condition are equal in anticipation accuracy, is twelve times more likely than the alternative hypothesis that these two conditions differ in button press accuracy (BF = 0.08). This provides "strong" evidence for the null hypothesis.

Comparing the subject means of the BIAS variable with the Bayesian paired *t*-test resulted in "substantial" evidence (BF = 0.1) for the null hypothesis. This analysis allows us to conclude that there is no statistically reliable difference between the BIAS in the Advanced-Knowledge and the Natural Turn condition. So the participants' button press accuracy with the natural turns was just as good as when they had advance information about the content of the turn. This finding suggests that participants are indeed able to anticipate a turn ending, and that they are using this strategy to predict when a turn is going to end.

The significant difference between Scrambled-Word-Order and Noise condition indicates that having access to words (even though they were in the wrong order) still allowed them to anticipate better than chance.

Although there was no significant difference in the button press accuracy between the Advance-Knowledge and the Natural-Turn condition, the participants could still have reacted to signals to a certain extent. If the participants used both anticipation and reaction as a strategy this should result in a lower *response consistency*. To investigate the response consistency over conditions we computed the Entropy for every stimulus/condition pair (Shannon, 1948). The Shannon Entropy is a measure of uncertainty: the more the responses are distributed over different intervals the higher the Entropy. If the participants used only one strategy to estimate when the turn is over, the Entropy should be lower. However, if the participants used both reaction and anticipation, their responses should be more highly distributed, resulting in a higher Entropy.

In **Figure 4** the average Shannon Entropy (using a bin-width of 250 ms; see De Ruiter et al., 2006 for details) is shown for every condition. We can only show a by-item analysis as these Entropy values can only be meaningfully computed for individual stimuli over entire response distributions.

As in the BIAS analysis, an ANOVA of the Entropy showed a main effect for condition $F2(3,297) = 62.5$, $p < 0.001$).

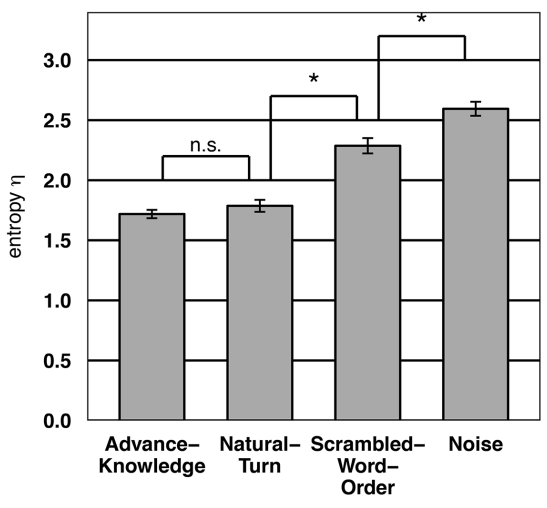


FIGURE 4 | Average Shannon Entropy per stimulus/condition as measured in Experiment 1. Asterisk indicates statistical significance at the 0.05 level.

Bonferroni-corrected paired *t*-tests revealed that all differences between individual conditions were significant ($p < 0.001$), the exception being the difference between Advance-Knowledge and Natural-Turn.

Again we compared the Entropy values in the Advance-Knowledge and Natural-Turn condition using a Bayesian paired *t*-test. The analysis (BF = 0.2) provided "substantial" evidence for the null hypothesis (no difference between Advance-Knowledge and Natural-Turn in button press consistency).

The analysis of the participants' button press consistency supports the interpretation of the BIAS results. The results showed that the Entropy in the Natural-Turn condition and the Advance-Knowledge condition was comparable. Thus, in the Natural-Turn condition the participants were able to consistently and accurately anticipate the turn-end and consequently used anticipation as a strategy to tell when a turn was over.

In contrast, in the Scrambled-Word and the Noise condition the Entropy values were significantly higher than in the other two conditions. This suggests that the participants tried to anticipate the turn-end rather than just waited for the end of the fragment, which lead to significantly broader distributed responses. In addition, the average Entropy in the Scrambled-Word order condition was significantly lower than in the Noise condition. This corresponds to the BIAS analysis above where participants in the Scrambled-Words condition were significantly more accurate in detecting the turn end. Hence, the participants are more consistent and accurate in the end-of-turn projection when they have access to words compared to when they only hear noise. One explanation of this finding could be that even with the scrambled word order listeners are able to recognize the basic meaning of the turn, enabling them to roughly guess when the turn finishes. Additionally, it is possible that once the participants "gambled" that a certain word was the last word, they could anticipate the end of that word, as suggested by research on auditory word recognitions (Marslen-Wilson and Welsh, 1978; McClelland and Elman, 1986; Marslen-Wilson, 1987).

We showed in Experiment 1 that listeners in dialog are indeed able to anticipate the end of the speaker's turn and that they consistently use this ability to predict when a turn is going to end. When listening to the natural turns the participants showed the same response accuracy and consistency as when they knew the end of the turn in advance. Our results are in line with earlier findings that listeners anticipate turn endings and that natural language is predictable to a certain degree (De Ruiter et al., 2006; Magyari and De Ruiter, 2012; Casillas and Frank, 2013; Keitel et al., 2013; Magyari et al., 2014). Hence, in the first experiment we were able to show that anticipation is the primary mechanism underlying smooth turn-taking, and that participants consistently use this strategy to detect a turn ending. Thus, our results support the turn-taking model proposed by Sacks et al. (1974). Nevertheless, reaction to the turn end might well serve as some kind of a "backup" mechanism in the case when the anticipation of the turn ending is, for whatever reason, not possible.

We now have an empirically derived distribution of anticipation times, from a task in which the participants were asked

to anticipate turn-ends, and had the information to do so. To find out about the distributional properties of the reaction process, which we assume also plays a role, we need to study the reaction time distribution of participants that had no information to anticipate (as in the Noise condition of Experiment 1) but in addition, were not instructed to anticipate, but rather to respond to the end of the stimulus. To this end, we conducted Experiment 2.

EXPERIMENT 2

Heldner and Edlund (2010) suggested that turn transitions with a gap longer than 200 ms are potentially explainable by assuming that participants respond to signals at the end of the turn. As we discussed in the introduction, this assumption does not capture the stochastic nature of the time course of the two processes involved. Instead, we assume that distributions of natural floor transfer are actually a stochastic mixture of an anticipation and a reaction time distribution. We wanted to empirically estimate the distribution of reacted responses in order to be able to estimate the proportion of turn-transitions that we were reasonably sure were reactions (and not to anticipations) to turn transitions.

An empirically estimated anticipation distribution is provided by the Advanced-Knowledge condition of Experiment 1. In Experiment 2 we want to find the other distribution based on pure reaction time. To this end, we used the Noise stimuli from Experiment 1, but now explicitly instructed the participants to respond only after they perceived the end of the fragment.

MATERIALS AND METHODS

Participants

Twenty native speakers of German participated in the second experiment (14 females, 6 males). None of the participants in Experiment 2 had taken part in Experiment 1.

Stimuli and design

Each participant was presented with all of the 100 noise target stimuli created in Experiment 1. In addition we took four stimuli from the practice block for practice purpose. There were two experimental lists, whereas in the second list the presentation order of the stimuli (including the practice trials) was reversed.

Procedure

The participants received a written instruction that they had to listen to noise fragments and press a button as soon as the noise stopped.

Within the experiment the participants were presented first with the four practice trials followed by the 100 target stimuli. After the practice trials the participants got the opportunity to ask questions. After the presentation of the first 50 target stimuli there was a break. The participants had to start the presentation of the remaining 50 stimuli by pressing a button, so that they could determine the length of the break by themselves. The participants were randomly assigned to one of the two experimental lists (10 per list).

Results and discussion

The reaction time distribution obtained in this experiment is presented together with the anticipation distribution from Experiment 1 in **Figure 5**. As expected, the reaction time distribution shows a pronounced sharp peak at a positive FTO (i.e., after the stimulus) whereas the anticipation distribution is broader and extends into the negative FTO range. In addition, the mode of the anticipation distribution is at a negative FTO value. The intersection of the two distributions characterizing the response time at which anticipation and reaction are equally probable is in good agreement with the 200 ms cut-off value proposed by Heldner and Edlund (2010). Nevertheless, the broad overlap of the two distributions shows clearly that the use of a categorical cut-off would not do justice to the stochastic nature of these two processes.

This is why we define a new measure designed to capture the *relative probability* of anticipation and reaction. The so-called Reaction Anticipation Proportion (RAP) value is defined as the natural logarithm of the ratio of anticipation and reaction probability.

$$RAP(t) = \log_e \left(\frac{P_{Ant}(t)}{P_{Reac}(t)} \right) \quad (1)$$

Equation 1: Definition of the RAP value as logarithmized ratio of the anticipation $P_{Ant}(t)$ and reaction probabilities $P_{Reac}(t)$ at time t.

In Eq. (1) $P_{Ant}(t)$ and $P_{Reac}(t)$ denote the probability that a response at time t was an anticipation or reaction, respectively. These probabilities were computed in R (R Development Core Team, 2009) using the density distributions (cosine kernel and 2.5 Sheather and Jones (1991) bandwidth) from the Advanced-Knowledge condition of Experiment 1 and the Noise condition of Experiment 2. To account for noise in the data leading to possibly infinite RAP values we used a cutoff value of 10^{-4} in the factor calculations. Due to the log-scale of the RAP ratio negative values corresponds to a higher probability of reaction whereas a positive value indicates that anticipation is more likely.

The RAP as a function of FTO is presented in **Figure 6**. The RAP is positive for a broad FTO interval ranging from about −750–200 ms and negative for FTO values in the interval from about 200–550 ms. Hence, reaction is more probable only in a relatively brief time interval. In addition, the influence of the pronounced sharp peak of the reaction distribution on the RAP value is weakened by the non-vanishing anticipation probability in the corresponding FTO range.

To demonstrate and validate the use of the RAP measure we applied it to the data analysis of the Natural turn and Noise conditions of the first experiment. The mean RAP value of the Natural turn condition was 0.60 and of the Noise condition −0.53. This supports our interpretation of the results of Experiment 1 that in the Natural turn condition the participants anticipated the end of the turn. In contrast, the responses in the Noise condition were predominantly based on reaction. It is noteworthy that the absolute value of the mean RAP of the two conditions are comparable, indicating that anticipation and reaction are about equally probable in the corresponding conditions.

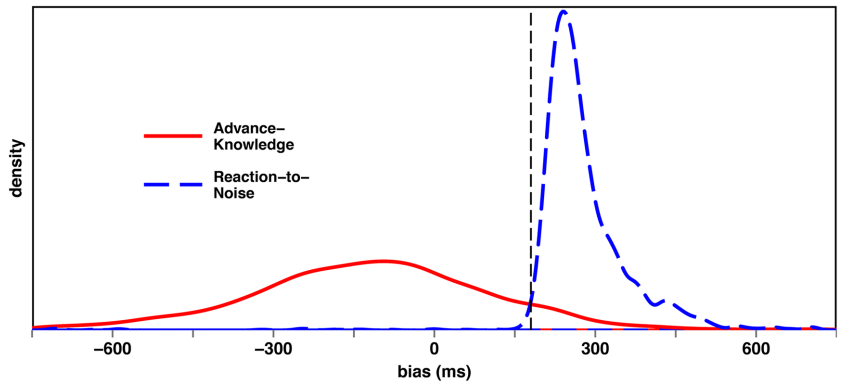


FIGURE 5 | Anticipation and reaction intervals.

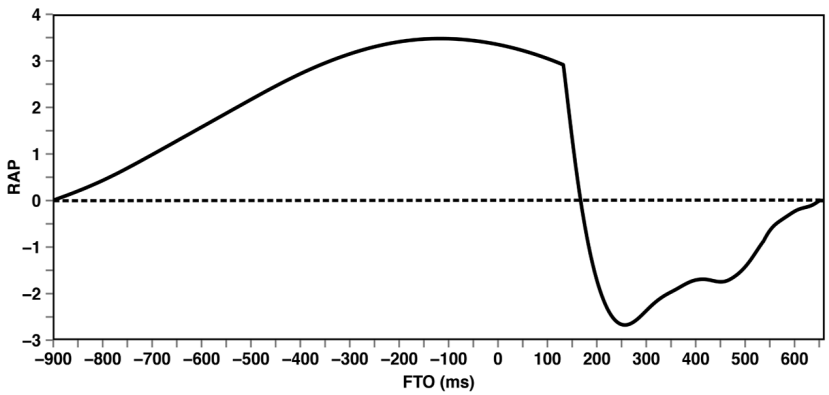


FIGURE 6 | Reaction Anticipation Proportion (RAP) value as a function of the FTO value.

We used the RAP measure to study the relative contribution of semantic and syntactic information in end of turn anticipation in Experiment 3.

EXPERIMENT 3

In Experiment 3 we investigated the relative role of syntax and semantics as a cue for end-of-turn anticipation. Experimental as well as corpus-based studies (Grosjean and Hirt, 1996; Selting, 1996; Caspers, 2003; De Ruiter et al., 2006; Magyari and De Ruiter, 2012; Keitel et al., 2013; Magyari et al., 2014) suggest that lexico-syntactic information serves as the main information source for end-of-turn prediction. But in these studies semantic and syntactic information is confounded, so the relative role of the individual source of information in turn anticipation cannot be established.

To tease apart semantic and syntactic information in natural communication we used the widely used distinction between open and closed class words in linguistics and psycholinguistics. Open class words in German contain nouns, verbs, adjectives, adverbs, and proper names. The open class words are "rich in referential meaning" (Chiarello and Nuding, 1987, p. 539) and "are the main bearers of meaning in language, providing the building blocks for the overall sense that is contained in spoken and written sentence" (Brown et al., 1999, p. 261). New words are easily added to the item set and they constitute the main part of our vocabulary (Segalowitz and Lane, 2000). The closed class category in German contains prepositions, articles, conjunctions, pronouns, modal verbs, quantifiers, and particles. Closed class words are semantically empty and serve primary a syntactic role (Crystal, 1988, p. 37). They serve to build the "structural skeleton of the sentence" (Kedar et al., 2006, p.325) and bear solely grammatical information (Jakubowicz and Goldblum, 1995, p. 247). The closed class contains a specified set of items, in which the addition of new objects trough cultural change is very slow (Segalowitz and Lane, 2000). Although closed class words only form a minority of our vocabulary, they are used much more frequently than open class words (Baayen et al., 1995; Rochon et al., 2000). To summarize, "the distinction between open- and closed class words can be seen as a basic reflection of the separation between syntax and semantics" (Brown et al., 1999, p. 261)[2].

[2]We are aware that open class words can contain syntactic cues, such as inflections and agreement on verbs and case marking on nouns, and closed class words can also contain referential information. Particularly in the Closed-Class-Words condition information as inflections could have helped to better anticipate the end of the turn.

Therefore, in this experiment, we operationalized semantic information as open class words, and syntactic information as closed class words. To address the question which information source listeners use for anticipation, we conducted a similar reaction time experiment as in Experiment 1. We manipulated the presence of semantic and syntactic information in the turn fragments from Experiment 1 by acoustically manipulating the recognizability of open- and closed class words. To evaluate the influence of these manipulations on the anticipation and reaction probability we used the RAP measure introduced before.

If only syntax is used for end-of-turn prediction, then the absence of closed class words should result in a decrease in anticipated and an increase of reacted responses. On the other hand, if semantic information is used for the anticipation of a turn ending, we expect a deteriorated anticipation performance in absence of open class words. However, if both semantic and syntactic information are used to the same extent, then the effect should be similar in absence of content as well as closed class words.

MATERIALS AND METHODS

Participants

Eighty native speakers of German who neither participated in Experiment 1 nor in Experiment 2 participated in Experiment 3 (53 females, 27 males).

Stimuli

The same 100 target and 16 practice turns as in Experiment 1 were used. We created four different versions of each turn fragment (see **Table 4** for an example of one experimental stimuli in all conditions). Natural Turn: the target turn was presented as it occurred in the natural conversation. Closed-Class-Words-Removed: the closed class words were "removed" by low-pass filtering (at 500 Hz Hanning Window). Open-Class-Words-Removed: by low-pass filtering we "removed" the open class words (at 500 Hz Hanning Window). Intonation-Only: the whole turn was low-pass filtered (at 500 Hz Hanning Window) so that no words could be recognized, but intonation remained intact. This condition served as a comparative baseline since neither syntactic nor semantic information were left in the turn fragment.

In the conditions Open-Class-Words-Removed and Closed-Class-Words-Removed the number of filtered words were made equal to the minimum number of open class words and closed class words in the turn. In this way we made sure that the number of filtered open and closed class words were the same for the same source stimulus. The decision which words were low-pass filtered was randomized. In order to control for subjective loudness between conditions and stimuli we again adjusted the loudness of all stimuli to a reference sone value.

Design

Each participant in the experiment was presented with three trial blocks: (1) Natural-Turn, (2) Intonation-Only, (3) The stimuli from the Closed-Class-Words-Removed and Open-Class-Words-Removed condition. The latter were presented within one block. The blocks Natural-Turn and Intonation-Only contained 25 and the combined block Closed-Class-Words-Removed and Open-Class-Words-Removed 50 target turns (25 stimuli from Closed-Class-Words-Removed and 25 stimuli from Open-Class-Words-Removed). Within each block there were four practice trials followed by 25 and 50 target turns, respectively. We created eight experimental lists. As in Experiment 1, we permutated the block order in four of these lists according to a Latin-square design. The remaining four lists were the same as the first four lists with the presentation order of the target stimuli and the practice trials reversed. As in Experiment 1 the lists were constructed so that all of the 100 target stimuli appeared in all four conditions across the lists but none appeared twice within the same experimental list.

Procedure

We used the same procedure as in Experiment 1.

Results and discussion

Figure 7 shows the response distributions for the four different conditions. **Figure 8** shows the average RAP values for the different conditions. The average positive RAP values in the Natural-Turn and the Closed-Class-Words-Removed condition indicate that the participants anticipated more frequently than reacted to the end of the turn in these conditions. In the Open-Class-Words-Removed and the Intonation-Only condition the participants reacted more often to the end of the turn.

An ANOVA on the RAP values showed a significant effect for presentation condition (by subjects: $F1(3,315) = 47.85$, $p < 0.001$, by items: $F2(3,297) = 74.11$, $p < 0.001$). Bonferroni-corrected paired t-tests revealed that all differences between individual conditions were significant ($p < 0.001$).

The BIAS distributions of the critical conditions Closed-Class-Words-Removed and Open-Class-Words-Removed, shown in **Figure 9**, supported the RAP analysis.

The results showed that when closed class words are removed participants are still able to anticipate the turn ending, although compared to the Natural turn condition the anticipation performance deteriorated. But when the participants could only identify closed class words (and not open class words) they reacted significantly more frequently to the turn end than when only open class words were identifiably.

Table 4 | Example of one experimental turn in all four conditions (underlined the respective low-pass filtered words).

Condition	Example
Natural-Turn	ich äh warte erstmal auf meine schwester und rufe die dann heute an
Closed-Class-Words-Removed	ich äh warte erstmal auf meine schwester und rufe die dann heute an
Open-Class-Words-Removed	ich äh warte erstmal auf meine schwester und rufe die dann heute an
Intonation-Only	ich äh warte erstmal auf meine schwester und rufe die dann heute an

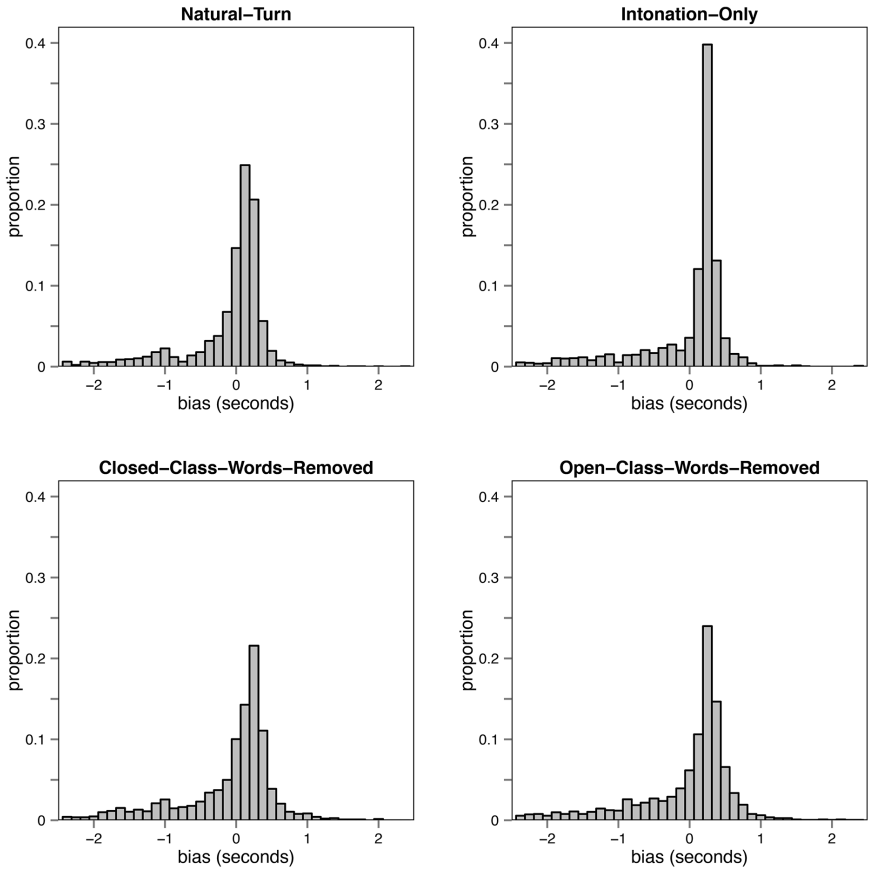


FIGURE 7 | Response distributions per condition from Experiment 3.

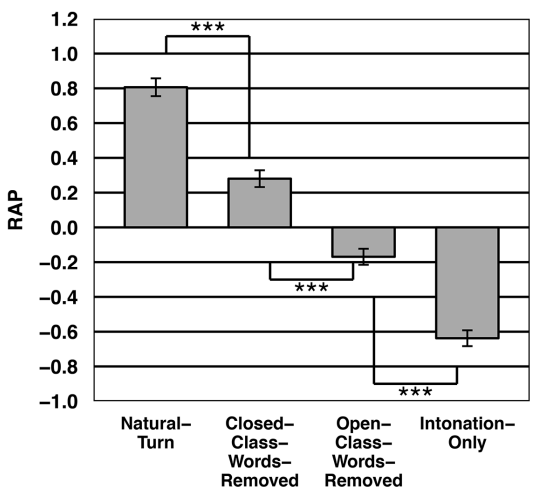


FIGURE 8 | Average RAP value per condition from Experiment 3. Asterisks indicate statistical significance at the 0.001 level.

The results suggest that semantic information is more important than syntactic information for end-of-turn anticipation. If no semantic information is available, it is less likely for the listeners to anticipate the turn ending. This also means that only knowing the syntactic frame and number of slots is not sufficient to estimate the timing of the turn. Nevertheless, the anticipation performance increased significantly when both open class words as well as closed class words were available. This could be explained by the fact that by removing closed class words the prediction of the content of the turn is also hampered. So for maximal anticipation performance listeners need semantic as well as syntactic information, probably because they need to be able to project the content of the turn. These results support the findings of Magyari and De Ruiter (2012) and Magyari et al. (2014) that listeners project the content of the turn to be able to estimate its duration.

Another interesting finding is that participants' anticipation performance was significantly higher when they got only closed class words compared to when only intonational and rhythmical information was available. This indicates that on top of the prosodic properties the syntactic structure provides additional anticipation cues.

Taken together these results suggest that semantic information is the most essential cue for anticipation. But to be maximally capable to anticipate a turn ending listeners need both semantic and syntactic information, since only the combination of both information sources allows for a correct projection of the content of the turn.

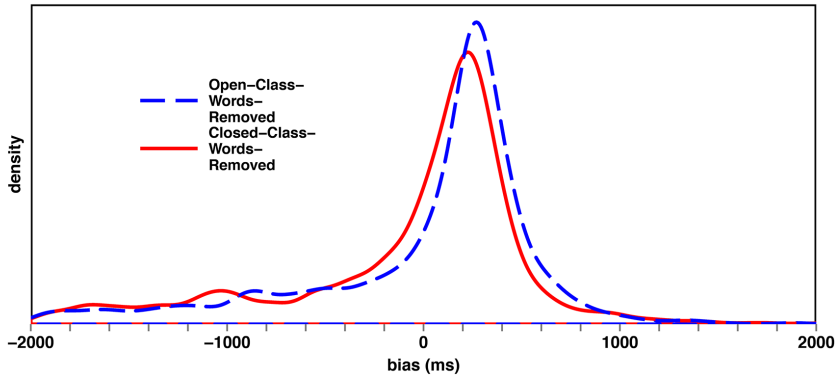


FIGURE 9 | Density plot of the BIAS distribution (in ms) of the Open-Class-Words-Removed and the Closed-Class-Words-Removed conditions. A paired *t*-test revealed that the difference of the means of the two conditions was significant [$t(99) = -3.23$, $p < 0.01$].

GENERAL DISCUSSION

In this study we addressed three different questions. First we investigated whether the observed accuracy in natural turn-taking was primarily due to anticipation or reaction to signals. Second, we wanted to quantitatively estimate the relative contribution of anticipation and reaction processes to the observed distribution of floor transfer timings. Finally, we studied the relative contribution of semantic and syntax in the timing of turn transitions.

In Experiment 1 we conducted a button press experiment in which we manipulated the information necessary for anticipation. The results showed that the listeners' response accuracy and consistency are similar when they (a) heard the natural turn and (b) when they are maximally able to anticipate the turn end by having advance information about the turn content. We conclude that listeners are indeed able to anticipate a turn ending and that they use this strategy consistently to tell when a turn is going to end. Thus, our findings support the functioning of the turn-taking mechanism proposed by Sacks et al. (1974). But it appears plausible that reaction to the turn ending could function as a "backup" mechanism in case of failures to anticipate turn-endings timely.

The data collected in Experiment 1 allowed us to estimate an empirical distribution for pure anticipation, so we proceeded to assess the counterpart distribution for pure reaction in Experiment 2 by explicitly instructing participants to react to the end of noise signals. We combined the two distributions to estimate the RAP, which represents the relative probability for a turn transition to have been guided by anticipation or reaction. By instructing the participants to react to the offset of a noise signal we estimated the 'other extreme' of anticipation, namely responding to the very end of a stimulus. We assume that a reaction to the offset of noise and a reaction to possible signals occurring at the very end of the turns (such as intonational patterns occurring immediately before the end of turns) are comparable from a reaction time point of view. It should be pointed out that it is also possible that conversationalists react to signals that occur before the very end of the turn. Because in our approach we assessed only the extreme opposites of pure (in the sense of 'as pure as practically achievable') anticipation and reaction, our data do not allow for an estimation of the possible contribution of such responses to the relative proportion of anticipation and reaction.

In Experiment 3 we investigated the effect of the presence or absence of semantic and syntactic information on the anticipation and reaction probability using the RAP measure. The results showed that the participants were still able to anticipate the end of the turn when they got access to semantic information. With only syntactic information available, the participants started to rely more on reaction. However, we found that to be maximally able to anticipate, listeners need syntactic information as well as semantic information. The absence of syntactic information hampers the projection of the content of the turn. We concluded that for anticipation both semantic and syntactic information are needed. Nevertheless, it appears that semantic information is a more important cue than syntactic information.

The RAP measure introduced in Experiment 2 is not only an analysis tool for the characterization of turn transitions but implies an inherently *stochastic* view of the turn taking process. By empirically estimating, for the first time, separate probability distributions for anticipation and reaction processes in end-of-turn detection, we were able to estimate the relative probability for a turn transition to be caused by anticipation or reaction at a given FTO value. This differs from the approach by Heldner and Edlund (2010) who suggested that any FTO larger than 200 ms could plausibly be explained by reaction, while FTOs shorter than this threshold indicate anticipation. The latter approach does not allow for the realistic possibility that anticipation could have been late, or reaction could have been early. Our RAP measure provides this information and allows for a more realistic assessment of the individual role of anticipation and reaction in turn taking. In addition, our model makes it possible to address many open questions in turn taking research, especially regarding the mechanism itself and its robustness. Finally, a very exciting (though time-consuming) possibility is to derive RAP/FTO curves for different languages. The RAP could reflect differences in the timing of how different languages deliver discourse-relevant information. Here, morphosyntactic differences between languages, for instance languages with relatively free word order relying heavily on case

marking versus languages like English with relatively fixed word order and a lean case marking system, may be reflected in different RAP/FTO curves. Alternatively, very similar RAP/FTO curves may suggest the presence of universals in the delivery of information in natural dialog.

Despite the mentioned advantages of the RAP measure over the strict threshold value suggested by Heldner and Edlund (2010), the RAP measure also does not incorporate the possibility of an intentionally delayed turn, for instance when that turn constitutes a 'dispreferred response.' Although this can be shown to happen in natural conversations, it is a situation that is difficult to recreate in a button press experiment; in our experimental setting we instructed the participants either to press the button when they thought the turn finished, or when the sound fragment was over. In this situation, we could not give the participants an interactional reason to delay their responses.

We showed in Experiment 3 that semantic information is a more important cue for anticipation than syntax. This finding contradicts former studies (Sacks et al., 1974; Selting, 1996; Caspers, 2003; De Ruiter et al., 2006) which assume that listeners rely primarily on syntactic information for anticipation. But how could semantic information serve to enable listeners to anticipate the turn ending? One possibility is that listeners use semantic information to predict the content of the speaker's turn and thus are able to estimate which words will be produced to convey the content. This is in line with the findings of Magyari and De Ruiter (2012) and Magyari et al. (2014) that listeners are able to predict the upcoming words of a turn. Another possibility is that during their experience as conversationalists, listeners have over the years built up certain expectations about how much (new) semantic information, on average, a conversational turn tends to contain. If the amount of semantic information exceeds this expected amount, this could be exploited as a cue that the turn is about to end soon.

Another explanation for the importance of semantic information in turn-taking could be that in naturalistic contexts, the semantics may provide stricter constraints on the turn construction than syntax does. Syntax theoretically allows for an infinite extension of a turn by the addition of new constituents. Furthermore, non-sense sentences like the famous "Colorless beautiful green ideas sleep furiously" (Chomsky, 1957, p. 15) are syntactically correct but provide no reliable meaning to base anticipation on. In other words, the end of a "Jabberwocky" sentence is impossible to predict.

By presenting isolated turns from natural conversations and letting the participants respond to the end of the turn by a button press we could both keep the characteristics of natural speech and at the same time systematically manipulate the turn fragments in order to test our specific hypotheses. Nevertheless, by isolating the turns we are not able to consider the impact of dialog context on anticipation. The discourse context could add information about the speaker's illocutionary intentions in the turn that is being produced, which in turn could help the listener anticipate its content. It is an interesting issue for future research whether, and if so, how, the discourse context can improve the anticipation of a turn ending.

AUTHOR CONTRIBUTIONS

The authors contributed to the following activities. Carina Riest and Jan P. de Ruiter designed the studies. Carina Riest and Annett B. Jorschick collected the data. Carina Riest, Annett B. Jorschick, and Jan P. de Ruiter analyzed the data. The paper was drafted by Carina Riest and revised by Annett B. Jorschick and Jan P. de Ruiter. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

ACKNOWLEDGMENT

This work was part of project B5 of the SFB 673 "Alignment in Communication" funded by the Deutsche Forschungsgemeinschaft (DFG).

REFERENCES

Baayen, R. H., Piepenbrock, R., and Bulickers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Boersma, P., and Weenink, D. (2012). *Praat: Doing Phonetics by Computer. [Computer Software Manual*]. Available at: http://www.praat.org ([Computer program] Version 5.3.14)

Brown, C. M., Hagoort, P., and Ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: open-and closed-class words. *J. Cogn. Neurosci.* 11, 261–281. doi: 10.1162/089892999563382

Casillas, M., and Frank, M. C. (2013). "The development of predictive processes in children's discourse understanding," in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Science Society), 299–304.

Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *J. Phon.* 31, 251–276. doi: 10.1016/S0095-4470(03)00007-X

Chiarello, C., and Nuding, S. (1987). Visual field effects for processing content and function words. *Neuropsychologia* 25, 539–548. doi: 10.1016/0028-3932(87)90078-9

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Crystal, D. (1988). "Linguistic levels in Aphasia," in *Aphasia*, eds F. C. Rose, R. Whurr, and M. A. Wyke (London: Whurr), 23–45.

De Ruiter, J.-P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's Turn: a cognitive cornerstone of conversation. *Language (Baltim.)* 82, 515–535. doi: 10.1353/lan.2006.0130

Duncan, S. J. (1972). Some signals and rules for taking speaking turns in conversation. *J. Pers. Soc. Psychol.* 23, 283–292. doi: 10.1037/h0033031

Duncan, S. J. (1973). Toward a grammar for dyadic conversation. *Semiotica* 9, 29–46. doi: 10.1515/semi.1973.9.1.29

Duncan, S. J., and Fiske, D. W. (1977). *Face-to-Face Interaction: Research, Methods, and Theory*. New York: Erlbaum.

Duncan, S. J., and Niederehe, G. (1974). On signalling that it's your turn to speak. *J. Exp. Soc. Psychol.* 10, 234–247. doi: 10.1016/0022-1031(74)90070-5

Ford, C. E., and Thompson, S. A. (1996). "Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns," in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge, MA: Cambridge University Press), 134–184.

Goodwin, C. (1986). Between and within: alternative sequential treatments of continuers and assessments. *Hum. Stud.* 9, 205–217. doi: 10.1007/BF00148127

Gravano, A., and Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Comput. Speech Lang.* 25, 601–634. doi: 10.1016/j.csl.2010.10.003

Grosjean, F., and Hirt, C. (1996). Using prosody to predict the end of sentences in English and French: normal and brain-damaged subjects. *Lang. Cogn. Process.* 11, 107–134. doi: 10.1080/016909696387231

Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *J. Phon.* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002

Indefrey, P., and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–44. doi: 10.1016/j.cognition.2002.06.001

Jakubowicz, C., and Goldblum, M. C. (1995). Processing of number and gender inflections by French-speaking aphasics. *Brain Lang.* 51, 242–68. doi: 10.1006/brln.1995.1060

Jeffreys, H. (1961). *Theory of probability*. 3rd Edn. Oxford: Oxford University Press, Clarendon Press.

Jescheniak, J. D., Schriefers, H., and Hantsch, A. (2003). Utterance format effects phonological priming in the picture-word task: implications for models of phonological encoding in speech production. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 441–454. doi: 10.1037/0096-1523.29.2.441

Kass, R. E., and Raftery, A. E. (1995). Bayes Factors. *J. Am. Stat. Assoc.* 90, 773–795. doi: 10.1080/01621459.1995.10476572

Kedar, Y., Casasola, M., and Lust, B. (2006). Getting there faster: 18- and 24-month-old infants' use of function words to determine reference. *Child Dev.* 77, 325–38. doi: 10.1111/j.1467-8624.2006.00873.x

Keitel, A., Prinz, W., Friederici, A. D., von Hofsten, C., and Daum, M. M. (2013). Perception of conversations: the importance of semantics and intonation in children's development. *J. Exp. Child Psychol.* 116, 264–277. doi: 10.1016/j.jecp.2013.06.005

Kendrick, K. H., and Torreira, F. (2014). The timing and construction of preference: a quantitative study. *Discourse Process.* doi: 10.1080/0163853X.2014.955997

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., and Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *J. Cogn. Neurosci.* 26, 2530–2539. doi: 10.1162/jocn_a_00673

Magyari, L., and De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25, 71–102. doi: 10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cogn. Psychol.* 10, 29–63. doi: 10.1016/0010-0285(78)90018-X

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0

Morey, R. D., and Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychol. Methods* 16, 406–419. doi: 10.1037/a0024377

Morey, R. D., Rouder, J. N., and Jamil, T. (2014). *BayesFactor Package: Computation of Bayes Factors for Common Designs*. Available at: http://bayesfactorpcl.r-forge.r-project.org/.

O'Connell, D. C., Kowal, S., and Kaltenbacher, E. (1990). Turn-taking: a critical analysis of the research tradition. *J. Psycholinguist. Res.* 19, 345–373. doi: 10.1007/BF01068884

Power, R. J. D., and Dal Martello, M. F. (1986). Some criticisms of Sacks, Schegloff, and Jefferson on turn taking. *Semiotica* 58, 29–40. doi: 10.1515/semi.1986.58.1-2.29

R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing*. Available at: http://www.r-project.org

Rochon, E., Saffran, E. M., Berndt, R. S., and Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: further development and new data. *Brain Lang.* 72, 193–218. doi: 10.1006/brln.1999.2285

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *J. Math. Psychol.* 56, 356–374. doi: 10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* 16, 225–237. doi: 10.3758/PBR.16.2.225

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language (Baltim.)* 50, 696. doi: 10.2307/412243

Schegloff, E. A. (1982). "Discourse as an interactional achievment: some uses of 'uh huh' and other things that come between sentences," in *Roundtable on Languages and Linguistics: Analyzing Discourse: Text and Talk*, ed. D. Tannen (Washington: Georgetown University Press), 71–93.

Schneider, W., Eschman, A., and Zuccolotto, A. (2012a). *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools, Inc.

Schneider, W., Eschman, A., and Zuccolotto, A. (2012b). *E-Prime Users's Guide*. Pittsburgh: Psychology Software Tools, Inc.

Schnur, T. T., Costa, A., and Caramazza, A. (2006). Planning at the phonological level during sentence production. *J. Psycholinguist. Res.* 35, 189–213. doi: 10.1007/s10936-005-9011-6

Segalowitz, S. J., and Lane, K. C. (2000). Lexical access of function versus content words. *Brain Lang.* 75, 376–389. doi: 10.1006/brln.2000.2361

Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics* 6, 357–387.

Shannon, C. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Sheather, S., and Jones, M. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Ser. B* 53, 683–690.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Taylor, W. (1953). " Cloze procedure": a new tool for measuring readability. *Journal. Q.* 30, 415–433.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., and Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: an empirical comparison using 855 t tests. *Perspect. Psychol. Sci.* 6, 291–298. doi: 10.1177/1745691611406923

Yngve, V. H. (1970). "On getting a word in edgewise," in *Proceedings of the Sixth Regional Meeting of the Chicago Linguistic Society*, Chicago 6, 567–578.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 October 2014; accepted: 16 January 2015; published online: 02 February 2015.

*Citation: Riest C, Jorschick AB and de Ruiter JP (2015) Anticipation in turn-taking: mechanisms and information sources. Front. Psychol. **6**:89. doi: 10.3389/fpsyg.2015.00089*

This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 09 February 2015
doi: 10.3389/fpsyg.2015.00098

Unaddressed participants' gaze in multi-person interaction: optimizing recipiency

Judith Holler* and Kobin H. Kendrick

Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

Edited by:
Sonja A. E. Kotz, Max Planck Institute Leipzig, Germany

Reviewed by:
Leonhard Schilbach, University Hospital Cologne, Germany
Anne Keitel, University of Glasgow, UK

****Correspondence:***
Judith Holler, Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525XD, Netherlands
e-mail: judith.holler@mpi.nl

One of the most intriguing aspects of human communication is its turn-taking system. It requires the ability to process on-going turns at talk while planning the next, and to launch this next turn without considerable overlap or delay. Recent research has investigated the eye movements of observers of dialogs to gain insight into how we process turns at talk. More specifically, this research has focused on the extent to which we are able to anticipate the end of current and the beginning of next turns. At the same time, there has been a call for shifting experimental paradigms exploring social-cognitive processes away from passive observation toward on-line processing. Here, we present research that responds to this call by situating state-of-the-art technology for tracking interlocutors' eye movements within spontaneous, face-to-face conversation. Each conversation involved three native speakers of English. The analysis focused on question–response sequences involving just two of those participants, thus rendering the third momentarily unaddressed. Temporal analyses of the unaddressed participants' gaze shifts from current to next speaker revealed that unaddressed participants are able to anticipate next turns, and moreover, that they often shift their gaze toward the next speaker before the current turn ends. However, an analysis of the complex structure of turns at talk revealed that the planning of these gaze shifts virtually coincides with the points at which the turns first become recognizable as possibly complete. We argue that the timing of these eye movements is governed by an organizational principle whereby unaddressed participants shift their gaze at a point that appears interactionally most optimal: It provides unaddressed participants with access to much of the visual, bodily behavior that accompanies both the current speaker's *and* the next speaker's turn, and it allows them to display recipiency with regard to both speakers' turns.

Keywords: turn-taking, turn projection, eye gaze, eye-tracking, unaddressed participants

INTRODUCTION

The contrast formed by the white sclera surrounding a darker iris and pupil is unique to the human eye (Kobayashi and Kohshima, 2001). This contrast renders eye gaze a highly salient cue in interaction with others, and the pivotal role gaze plays in human communication has been demonstrated by numerous studies (see Argyle and Cook, 1976; Cook, 1977; Kleinke, 1986; Itier and Batty, 2009; Senju and Johnson, 2009; Rossano, 2012 for reviews). By now, we know a great deal about how gaze functions in dyadic encounters, such as to initiate interaction, signal address, receive addressee feedback, and coordinate turn transitions (e.g., Kendon, 1967, 1990; Argyle et al., 1973; Cary, 1978; Duncan et al., 1979; Goodwin, 1980; Bavelas et al., 2002; Lerner, 2003; Rossano et al., 2009). Here, we study gaze behavior with respect to another core aspect of social interaction, namely the precise timing of gaze and turns at talk in multi-person interaction. More precisely, we investigate how the cognitive processing of turns infuences gaze behavior of momentarily unaddressed participants during question–response sequences and consider the social opportunities this may create in a triadic conversation context.

THE TIMING OF TURNS AT TALK

In social interaction, a system of turn-taking organizes opportunities to speak. According to Sacks et al. (1974), turns at talk are constructed out of linguistic units that have recognizable structures, enabling a next speaker to project the structure in advance and, consequently, anticipate the possible completion of the unit. Subsequent research has examined the syntactic and prosodic structures that allow for the projection of a current turn and signal its possible completion (Ford and Thompson, 1996; Ford et al., 1996; Selting, 1996; Wells and Macfarlane, 1998; Auer, 2005; Local and Walker, 2012). Within the model, the first possible completion of such a unit constitutes a place, referred to as a transition-relevance place, at which a transition from current to next speaker may occur (Sacks et al., 1974; Selting, 2000). A set of rules and constraints in the model, such as a constraint on more than one speaker at a time (Sacks et al., 1974; Jefferson, 1986; Schegloff, 2000), accounts for the observation that transitions tend to occur with minimal overlap between turns. At the same time, rules, and constraints in the model lead to minimal gaps between turns. This is particularly remarkable since quantitative studies have shown that gaps between turns are most frequently

on the order of just 0–200 ms (Stivers et al., 2009; Heldner and Edlund, 2010). As Levinson (2013) has argued, short gaps between turns do not provide adequate time to prepare even a simple next turn, which psycholinguistic research has shown requires at least 600 ms (Indefrey and Levelt, 2004; Indefrey, 2011). This suggests that a next speaker must begin to plan the next turn well before the current one is complete, a psycholinguistic challenge in which projection of a current turn appears to play an important role (De Ruiter et al., 2006; Magyari and de Ruiter, 2012; Magyari et al., 2014).

THIRD-PERSON PERSPECTIVE EYE-TRACKING STUDIES ON TURN-TAKING

Recently, a new experimental paradigm has been developed for the study of the cognitive processes that underpin turn-taking from a third-person perspective. The general procedure involves participants being presented with a pre-recorded dialog or conversation between two people on a computer screen while their eye movements are tracked and timed with respect to the turns at talk they hear. Experimental studies using this novel paradigm have shed light on the precise timing of eye movements and turns at talk by measuring where observers of dialogs look and when they do so.

A study by Augusti et al. (2010) has shown that infants of just 6 months of age shift their gaze from current speaker to next speaker in accordance with the alternation of turns, thus, they argue, showing a sensitivity to the natural flow of conversation. Other studies have shown that, at least by 3 years of age, children are not only able to track who is speaking at any one time, but they are indeed able to anticipate upcoming turns, shifting their gaze to the next speaker often *before* he or she begins to speak (von Hofsten et al., 2009; Casillas and Frank, 2012, 2013; Keitel et al., 2013).

Studies using the same paradigm with adults have shown that they, too, tend to look reliably at the current speaker (Tice and Henetz, 2011; Casillas and Frank, 2012; Edlund et al., 2012; Hirvenkari et al., 2013). However, these studies have yielded discrepant findings regarding when observers begin to look to the next speaker. Foulsham et al. (2010) asked observers to watch a video of others performing a conversation-based group-decision task and to decide whom of these they would like to work with on a subsequent task. Their findings showed that observers fixated the next speaker on average 150 ms before they started to speak. Tice and Henetz (2011), Casillas and Frank (2012), and Keitel et al. (2013) measured the eye movements of observers of dialogs. Keitel et al. (2013) found that 54% of adults' gaze shifts occurred within a time window starting 500 ms prior to the end of the current turn and ending with the beginning of the next turn. The gaze shifts thus occurred while the current speaker was still talking, or during the gap between turns, providing clear evidence of anticipation of the next turn. Similarly, Tice and Henetz (2011) and Casillas and Frank (2012) found that the majority of their participants' eye movements to the next speaker occurred either during the gap between turns or within the first 200 ms of the next turn. Since it takes around 200 ms for a saccadic eye movement to be planned and launched (Salthouse and Ellis, 1980; Fischer and Ramsperger, 1984; Becker, 1991; Allopenna et al., 1998; Griffin and Bock, 2000), these gaze shifts must have been planned *prior* to the beginning of this next turn. Moreover, in at least some cases, observers shifted their gaze to the next speaker even before the *current* turn had ended (Casillas and Frank, 2012, 2013). Together, the findings from these studies suggest that observers of scripted dialogs and spontaneous group conversations engage in predictive cognitive processes that allow them to anticipate the beginnings of next turns, and, at least to some extent, also the ends of current turns.

However, two studies using truly spontaneous (rather than scripted or performed) dialogs have not found evidence for anticipatory looks to the next speaker. Edlund et al. (2012), too, have shown that observers track current speakers with their gaze, and although the precise timing of this gaze with respect to turn transitions is not provided, the data they do provide seem to suggest that looks to the next speaker before he or she started to speak were rare, if present at all. Hirvenkari et al. (2013), too, found that their observers looked at the next speaker only after he or she had already begun to speak. One possible reason for this, they state, could be that participants in other studies (e.g., Foulsham et al., 2010) may have been more eager to see the reactions of the participants due to the decision task they were asked to complete. They argue that the gaze behavior of their participants merely observing dialogs may have been "less tightly linked to the turn-taking than if the task would have been more engaging, or if the subjects would have actually taken part in the conversation" (Hirvenkari et al., 2013, p. 6). Thus, it is evident that the nature of the experimental task and the spontaneity of the conversational exchange may influence the temporal coupling of observers' eye movements and turns at talk. An investigation of the timing of eye movements and speaking turns while participants are engaged in actual conversation, processing spontaneous turns without them being required to complete an experimental task, is therefore an important next step.

METHODOLOGICAL CONSIDERATIONS

While there is some discrepancy in findings, studies using the novel third-person perspective eye-tracking paradigm described above have provided us with valuable first insights into how adults may process turns at talk and transitions between them, as well as how children acquire this skill during development. However, two issues emerge from this work.

The first issue has already been alluded to in the preceding section and concerns the third-person perspective as such. Recently, Schilbach (2010, 2014) and Schilbach et al. (2013) put forward a convincing argument for the urgency of a shift in experimental paradigm, stating that "recent conceptual and empirical developments consistently indicate the need for investigations that allow the study of real-time social encounters in a truly interactive manner. This suggestion is based on the premise that social cognition is fundamentally different when we are in interaction with others rather than merely observing them" (Schilbach et al., 2013, p. 393). Their argument, and the evidence they cite, concerns the abundance of paradigms in the field of cognitive neuroscience involving passive observation and the different insights interactive paradigms have provided in this domain. The latter immerse participants in 'online' social interaction rather than ask them

to observe offline interactions, thus creating reciprocal relations with sequences of actions and reactions shaping the communication between the participants (Wilms et al., 2010; Pfeiffer et al., 2013).

One important question that remains, therefore, is when participants shift their gaze from current to next speaker if they themselves are ratified participants in the conversation but momentarily unaddressed (Goffman, 1979, 1981; Clark and Carlson, 1982). If the degree of engagement that participants feel indeed influences their ability (or motivation) to project either current or next turns, then we might see more evidence of early gaze shifts when participants are directly immersed in a live conversation. An alternative possibility is, however, that the considerably reduced social context of third-person perspective paradigms underestimates the cognitive demands placed on processing turns at talk in spontaneous conversation. Participants may thus have *less* cognitive resources available for projection in live conversation, meaning gaze shifts may be primarily responsive to the next speaker beginning to speak rather than anticipatory. However, it could also be that eye movements in face-to-face interaction do not reflect the projection of current or next turns at all, but that the social norms and dynamics of conversation determine where participants look and when. Thus, while third-person perspective eye-tracking paradigms allow for a high degree of experimental control and manipulation to investigate eye movements during turn-taking, and the role semantics and prosody play in this context (Casillas and Frank, 2013; Keitel et al., 2013), they cannot necessarily tell us what guides participants' eye movements in more situated contexts such as spontaneous, multi-person interaction.

The second issue concerns the structure and analysis of turns at talk that have been used in third-person perspective eye-tracking studies. With the exception of some studies (Foulsham et al., 2010; Edlund et al., 2012; Hirvenkari et al., 2013), the stimuli in third-person perspective eye-tracking studies were scripted and strongly controlled, which has a range of implications. For one thing, it means that the dialogs were presumably fairly carefully spoken and had rather long gaps between turns and few if any overlaps. Indeed, in some cases the gaps between turns were 900 ms on average (Keitel et al., 2013), which considerably exceeds the ∼200 ms mean gap duration (Stivers et al., 2009) and the 0–200 ms mode of gap durations (Stivers et al., 2009; Heldner and Edlund, 2010) observed for spontaneous conversation. In fact, 900 ms gap durations are more representative of lengthened gaps marking dispreferred responses (Kendrick and Torreira, 2014). Careful pronunciation, lack of overlap, and relatively long gaps may, of course, all influence how turns are processed and projected.

Moreover, the structure of questions in spontaneous conversation is often complex, with more than one point of possible completion within a single turn. Conversation-analytic research on turn-taking has suggested that participants in conversation monitor turns for points at which they are recognizable as possibly complete because such points constitute opportunities for transition between speakers (Sacks et al., 1974). In the following example, the participant addressed by the question responds at the first point at which the question is possibly complete, even though the speaker of the question continues his or her turn, adding a term of address.

FD:IV:191 (Sacks et al., 1974, p. 702)
Desk: What is your last name [Loraine.
Caller: [Dinnis.
Desk: What?
Caller: Dinnis.

Thus, the first possible completion of a question is not necessarily the end of the turn, as a speaker can continue to speak past this point. If participants in conversation do indeed monitor turns at talk for points of possible completion, as Sacks et al. (1974) proposed, then we may find evidence for this in the eye movements of unaddressed participants in question–response sequences. However, in studies using the third-person perspective paradigm, either the turns used as stimuli were constructed to have simple structures in which the first possible completion of the turn was coterminous with its end, or multiple possible completions were not taken into account in the analysis. As a consequence, it is currently unknown how the gaze behavior of observers is timed with respect to points of possible completion prior to the ends of turns as such. The literature on third-person perspective eye-tracking paradigms has referred to eye movements that precede the end of a turn as anticipatory. Since first possible completions are often not the end of the turn, gaze shifts that are anticipatory with respect to the end of the turn may actually follow a first possible completion point, or may virtually coincide with this point. The extent to which eye-movements do or do not anticipate the possible completion of a turn matters for the interpretation of results from this paradigm within models of turn-taking behavior, thus further underlining the need for a systematic consideration of the intricate structure of turns.

THE PRESENT STUDY: INVESTIGATING EYE MOVEMENTS AND TURNS AT TALK *IN SITU*

The present study aims to shed light on the timing of eye movements and turns at talk by situating the third-person perspective eye-tracking paradigm within spontaneous, live conversations. To this end, using state-of-the-art technology, we studied a corpus of triadic conversations between friends and examined exchanges in which a speaker addressed a single participant, thus rendering the third a momentarily 'unaddressed participant' (Bolden, 2013; cf. 'unaddressed recipient,' Goffman, 1979, 1981; cf. 'side-participant,' Clark and Carlson, 1982; 'audience,' Levinson, 1988). More specifically, we tracked this person's eye movements during question–response sequences to measure whether, and if so at precisely which point, unaddressed participants moved their eyes from current to next speaker. This approach builds on earlier work by moving from scripted dialogs involving actors to natural multi-person interaction in which participants experience personal immediacy and co-presence, the turns at talk are of direct relevance to them, and participants may become the addressee at any given moment. Moreover, the measurements of turns and gaps between them are not determined by the experimenter or actors but are natural in content and length. Further, we not only

consider questions in their entirety but also apply a more fine-grained analysis, tackling the intricate structure of spontaneous questions by examining the timing of eye movements with respect to first possible completions, as well as the end of turns. Thus, we aim to answer not only the question of how eye movements are timed with respect to turns, but also to what extent they are governed by the projection of the current or next turn. Finally, while to date all reports have discussed observers' gaze behavior across turn transitions in terms of the cognitive processes that underpin turn-taking, the present study also aims to consider the nature of this phenomenon as a social behavior. This will help us understand whether we are dealing with a turn-taking phenomenon *per se* or with one that belongs to some other order of conversational organization.

MATERIALS AND METHODS

PARTICIPANTS AND CORPUS

The corpus consists of ten groups of participants engaging in casual conversations in English recorded at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. The recordings include both ten triadic (three participants) and ten dyadic (two of the three participants) conversations[1] (for an eye-tracking corpus of dyadic interactions in Flemish, see Brône and Oben, 2014). All conversations are ∼20 min. in length. For the eye-tracking analyses reported here, seven of the ten triadic conversations were analyzed as calibration was poor for the remaining three. All participants were native speakers of English recruited from the general Nijmegen population and knew each other prior to the recording session (except for one triad in which one person knew both of the other two participants who had not themselves met before). Their ages ranged from 19–68 years (Mean age = 30 year). Two of the conversations were all female, two all male, and three conversations consisted of two female and one male participant.

LABORATORY SET-UP AND TECHNICAL EQUIPMENT

The recordings took place in a sound proof room equipped with professional lighting suitable for high quality audio and video recording. Participants sat in standard height chairs with armrests, arranged in a triangle with the chairs equidistantly placed from one another. A ceiling microphone recorded the entire conversation. Each participant wore a head-mounted lightweight uni-directional microphone (Shure SM10A), which recorded only the respective participant's voice, and a pair of eye-tracking glasses (SMI, sampling rate 30 Hz). In addition, three HD video cameras (Canon Legria HFG10, 25 fps) recorded frontal views of each person (except for one triad where one of the three HD cameras failed to record; the respective participant's data was not included in the present analysis). Due to the spatial arrangement of the chairs with respect to the cameras, each person was also visible from the right and left side in the recordings made by the respective other two video cameras. **Figure 1** provides an overview of the laboratory set-up and equipment.

[1]Due to the present focus being on unaddressed participants' eye movements, the dyads are reported here for completeness sake but are not analyzed in the present study.

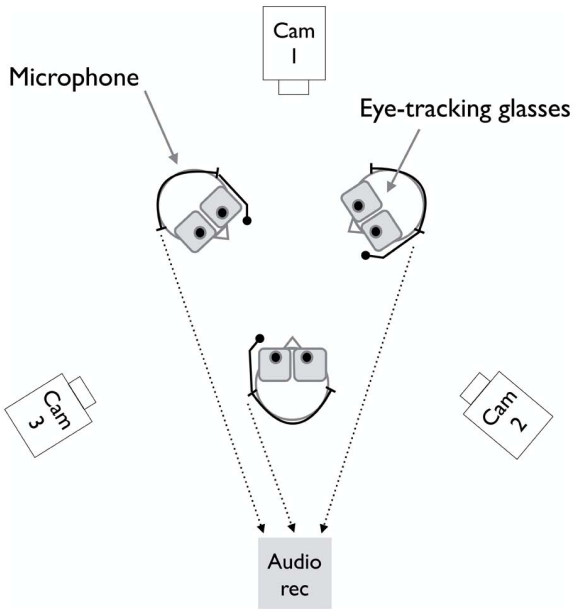


FIGURE 1 | Illustration of the technical laboratory set-up used in the present study.

For each session, the recorded material resulted in three individual videos from the cameras, three individual videos from the eye-trackers (exported from the SMI recording device with the gaze cursor overlaid onto the visual scene recorded by the video cameras of the eye-trackers), three individual audio files, and the audio file from the ceiling microphone. The audio tracks were recorded in sync using a four-channel audio recorder (Edirol/Roland R-44). The six video recordings and three individual audio recordings were combined and synchronized in Adobe Premier Pro CS4 and then exported as a single audio–video file for analysis (MP4) at 24 frames per second (see **Figure 2**). This resulted in a time resolution of approximately 41 ms, the duration of a single frame. The synchronization was based on audible and visible information from a clapperboard used at the beginning and end of each session.

PROCEDURE

Upon their arrival, participants were greeted by two investigators who conducted the study (JH and KK) and were handed study-packs, including information about the study and procedure of the session, forms asking about their language background, screening questionnaires ruling out motor and speech impairments, as well as consent forms and questionnaires about handedness and a variety of social dimensions. Once the study-packs had been completed by all participants (except for the social questionnaires, see below) and any queries had been answered, participants were seated in their chairs in the recording room. All equipment was prepared beforehand, allowing immediate fitting of the microphones and the eye-trackers (involving a three-point calibration procedure).

Each recording session lasted approximately 40 min. in total, with the first 20 min. constituting a trialogue phase and the

FIGURE 2 | Still frame from a synchronized six-video recording (one triad). Top panel shows the three eye-tracker videos including gaze cursor (in orange); bottom panel shows the three HD camera recordings. The video of each participant's view through the eye-tracker is positioned above the corresponding frontal HD video recording of this participant.

second 20 min. a dialog phase. Upon completion of the initial fitting procedure, the two investigators left the room and waited in an adjacent area until the first 20 min. had elapsed. At this point, they compared performance of the three eye-trackers and asked the person wearing the eye-tracker with the poorest calibration to leave the room. Once the remaining two participants had talked for another 20 min., all three were reunited in the recording room and asked to complete the social questionnaires contained in the study-packs. This was to ensure that questions about human communication and behavior (verbal and non-verbal) would not influence participants' behavior during the conversations. (The results from the social questionnaires are not of relevance for the present analysis and will not be discussed any further.) Participants were then asked one more time for their written informed consent relating to how their data should be handled, thanked, and financially compensated for their participation (26 euro per person). The entire test session lasted around 120 min. The study was approved by the Social Sciences Faculty Ethics Committee, Radboud University Nijmegen.

ANALYSIS

Question-response sequences

The present analysis focused on question–response (henceforth QR) sequences in which the question was addressed to a single participant who then produced a response. All QR sequences were identified by an experienced conversation analyst (Kobin H. Kendrick), resulting in a total of 281 questions and their responses (a subset of which was included in the final gaze shift analysis, see Eye Gaze). Criteria for identifying QR sequences in our dataset were based on the coding scheme proposed by Stivers and Enfield (2010, pp. 2621–2626). The precise beginnings and endings of the questions and the responses were determined in Praat 5.3.77 (Boersma and Weenink, 2014). In-breaths preceding responses were clearly audible in our recordings and were treated as the onset of the response ($N = 35$). In a small number of cases ($N = 2$) the response was exclusively non-verbal (e.g., head nods); in those cases the beginning of the response was timed to the first frame of visible movement. These annotations were then imported into ELAN 4.61 (Wittenburg et al., 2006).

Points of possible completion

All questions in the dataset were analyzed for the presence and location of points of possible completions before the end of the turn, drawing on conversation-analytic research on turn construction (Sacks et al., 1974; Ford and Thompson, 1996; Ford et al., 2002). For a point of possible completion to be identified, the turn at talk up to that point must have been hearable to the analyst as a possibly complete question in its context. This determination was made holistically, with attention to the syntax, prosody, and meaning of the question. For those questions with a point of possible completion before the end of the turn, the precise location of the first possible completion was annotated in ELAN. Crucially, the participants' gaze behavior was not considered in this analysis.

The analysis of points of possible completion revealed a number of recurrent types. If a turn contained two complete questions, a point of possible completion – represented here by a vertical bar – was identified after the completion of the first, whether the two questions were produced one after the other (e.g., "where does she

go? | where- where does she- what uni's she from?") or with a short silence between the two (e.g., "but is it good? | (0.1) or is it just (0.2) any money is good?"). If a turn contained a possibly complete question together with an increment, a contingent addition to a turn that continues its grammatical structure (Schegloff, 1996; Couper-Kuhlen and Ono, 2007), a point of possible completion was identified before the increment (e.g., "how are you finding it by the way"; "were you on a bike | at that time?"). If the turn contained a tag question, a frequent occurrence in the dataset, a point of possible completion was identified before the tag (e.g., "there was like a fifth one | wasn't there?; "you were at it too | right?"). And if the turn contained a possibly complete question followed by a turn component that could not have been projected or anticipated in advance, a point of possible completion was identified after the question (e.g., "all your family's in England | I expect?"; "so it's on campus this place?").

Interrater agreement between two coders (KK and JH) who independently identified the presence and precise location of points of possible completion in the dataset revealed strong reliability, $K = 0.72$ (85.7% agreement; Landis and Koch, 1977).

Eye gaze

The ELAN files containing the QR annotations were linked to the synchronized videos in order to annotate the unaddressed participants' eye movements during the QR sequences. These annotations were done manually, on a frame-by-frame basis. At each frame during the QR sequence, the gaze fixation point generated by the SMI software for the unaddressed participant (henceforth referred to as the gaze cursor) was categorized as being (1) on speaker A, (2) on speaker B, (3) on self (e.g., when looking at his or her own hands), (4) on the surroundings (e.g., the walls, the door, any equipment items in the room), or (5) as not identifiable from the eye-tracker data (i.e., the eye-tracker cursor was not visible in the respective video frames). Based on this coding scheme, 45 of the originally 281 QR sequences (16.0%) were discarded from further analysis of the unaddressed participants' eye movements due to insufficient data. (Note that the eye movement data of unaddressed participants is associated with considerably more data loss than the eye movement data for speaker A and speaker B. This is because, in our set-up, unaddressed participants often move their heads as well as shift their gaze to look from the current to the next speaker, and these movements tended to be performed quite fast and with the eyes being closed during the shift, thus obscuring the corneal reflection the eye-tracker needs to capture).

Out of the remaining 235 QR sequences, unaddressed participants moved their gaze from speaker A to speaker B in 105 (45.5%) QR sequences. In order to be considered a valid gaze shift for our analysis, the trajectory had to be one that clearly moved from A to B, without the gaze pausing elsewhere in between (such as on self or background objects). In the remaining 131 sequences, unaddressed participants either did not shift their gaze at all and instead fixated speaker A, speaker B, themselves, or the surroundings throughout, or they did move their eyes but in the opposite direction, that is, from speaker B to speaker A. While these cases are interesting in themselves, they tap into a different phenomenon than the one under investigation here and require analysis and discussion elsewhere.

Regarding those 105 QR sequences that did reveal a gaze shift from speaker A to speaker B (i.e., our final QR dataset), the average question duration was 2018 ms (Median = 1681 ms; minimum value = 328 ms; maximum value = 7667 ms), and the average response duration was 1899 ms (Median = 1312 ms; minimum value = 164 ms; maximum value = 8118 ms). Due to the highly dynamic nature of conversation brought about by, amongst other things, differences in personality, age, gender, closeness of friendship, and topic of discussion, the seven triads of course differed in the number of QR sequences they contributed to our analysis (they contributed 2, 8, 10, 12, 16, 22, and 35 QR sequences, respectively). Likewise, participants within the triads differed in the extent to which they contributed to the conversation by asking questions, but none of the conversations excluded participants (and those that asked fewer questions may, of course, have contributed more to the conversation in other ways, such as through tellings, jokes, responses, and so forth). Basing analyses of QR sequences in conversation on samples that are determined by the participants' spontaneous behavior, thus resulting in varying numbers of QR sequences across separate interactions, is the standard procedure for corpus studies and in line with existing research (e.g., Stivers et al., 2009, 2010; Gardner, 2010; Strömbergsson et al., 2013).

For these 105 QR sequences that did reveal a shift of the unaddressed participant's gaze from speaker A to speaker B, we identified *when* exactly this gaze shift occurred. The time window we took into consideration for identifying gaze shifts relevant for this analysis stretched from the beginning of A's turn to the end of B's turn. In all cases of gaze shifting from speaker A to speaker B within this time window, unaddressed participants looked at the face of speaker A and then moved their gaze from there to the face of speaker B. Using the frame-by-frame gaze annotations, we identified the first frame at which the gaze cursor left speaker A, defined as the frame at which the gaze cursor was no longer on, overlapping with, or directly adjacent to speaker A's head or technical head-gear (see **Figure 3**). At what time point before or during B's turn the unaddressed participant's gaze arrived at speaker B was not of relevance for the present analysis. Annotations were made in ELAN to measure the duration from the first gaze shift away from speaker A by the unaddressed participant to two points within the question turn: (1) the end of the turn and (2) the first point of possible completion of the question, for those questions that had a possible completion before the end of the turn. In addition to measuring the duration of these intervals (in ms), the values were set as either positive or negative. This was done to identify the temporal order of the respective events, with negative values indicating an anticipatory gaze shift before a point of possible completion or the end of a turn and positive values indicating the inverse.

However, we need to consider that it takes time to plan and launch these eye movements before they are observable. This process is estimated to take on average around 200 ms (Salthouse and Ellis, 1980; Fischer and Ramsperger, 1984; Becker, 1991; Allopenna et al., 1998; Griffin and Bock, 2000). We therefore calculated a value for the beginning of the assumed planning phase for each observed value by subtracting 200 ms.

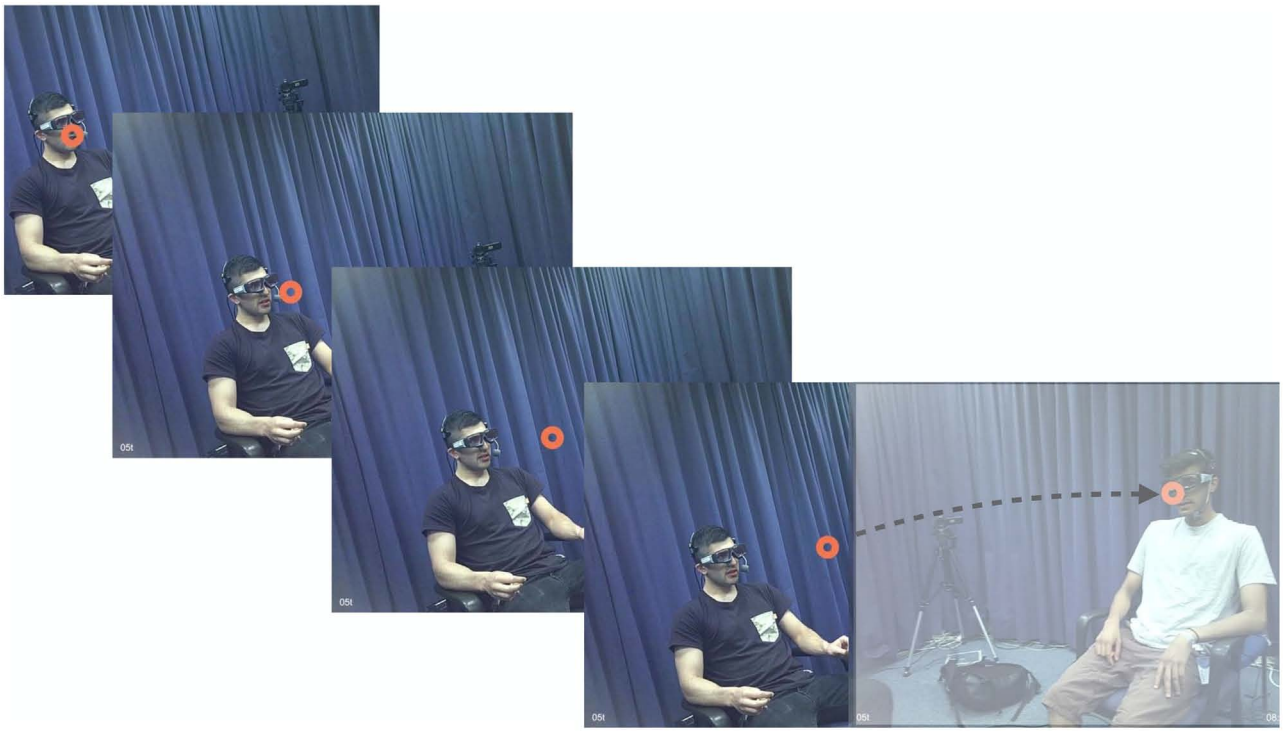


FIGURE 3 | Consecutive still images providing an example of an unaddressed participant's gaze shifting away from speaker A toward speaker B during a QR sequence. Frames 1 and 2 capture instances of the gaze cursor being classed as on speaker A (see coding criteria), whereas frame 3 captures the first gaze shift away from speaker A (moving to speaker B, frame 4).

Gaze coding was performed by two independent coders (LD and MvdG) blind to the study's predictions and assumptions. In addition, their coding was checked by one of the two senior analysts (JH and KK), and any errors in coding (of which there were remarkably few due to the clear categorical distinctions between gaze locations) were discussed and corrected. Due to the considerably more objective coding criteria applied for our gaze analysis in comparison to the identification of points of possible completion, formal reliabilities were calculated for the latter only.

Statistical analysis

All statistical analyses were conducted in R 3.1.1 (R Core Team, 2012). The density plots displayed in the Results section were generated using the Lattice package (Sarkar, 2008) with default kernel density estimation (Gaussian). Since these distributions render a smoothed curve (rather than a histogram) and an *estimate* of the mode, all mode values given should be considered close approximations of the true value and decimal places are not stated for those values. Note also that these distributions are based on binned data brought about by our video frame rate providing a measure every 41.7 ms (24 fps).

RESULTS

Out of the 105 QR sequences analyzed here, 54.3% ($N = 57$) of the questions had at least one possible completion before the end of the turn. Here, we focus our analyses first on the end of the turn and then on its first possible completion (which corresponded to the end of the turn for 45.7% of questions).

EYE MOVEMENTS TIMED WITH RESPECT TO THE END OF TURNS

First, we measured the time point of each first observed gaze shift away from speaker A (and toward speaker B) with respect to the end of speaker A's turn. This showed that the estimate of the mode of these data is located very close to the end of the question, namely just 50 ms prior to turn end (see **Figure 4**, solid line). Because, as we have already noted, experimental research indicates that overt eye movements are planned about 200 ms in advance of them being observed, the covert initiation of unaddressed participants' eye movements from speaker A to speaker B occurred most frequently around 250 ms prior to the end of questions (see **Figure 4**, dashed line).

On the whole, 60.0% ($N = 63$) of QR sequences were associated with observable gaze shifts that anticipated the end of the question turn. When taking into account the time it takes to prepare these gaze shifts, the percentage of anticipatory gaze shifts increases to 73.3% ($N = 77$).

EYE MOVEMENTS TIMED WITH RESPECT TO THE FIRST POSSIBLE COMPLETION

Because many of the questions in our data had a point of possible completion prior to turn end (as seen above), we carried out a second analysis in which we timed unaddressed participants' first gaze shift away from speaker A with respect to the

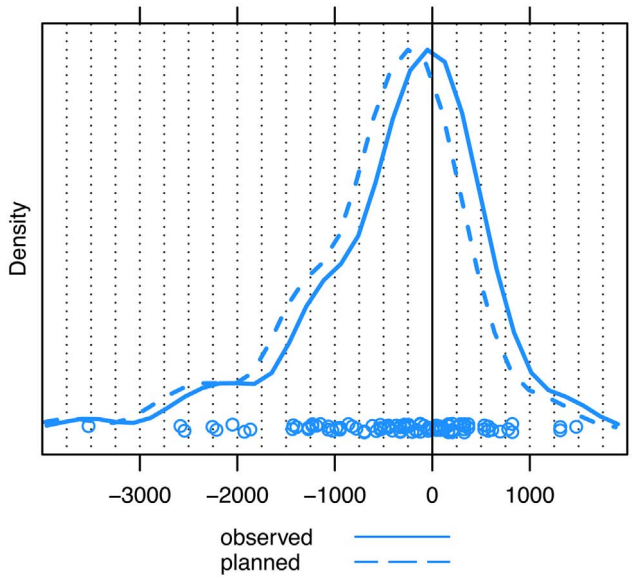


FIGURE 4 | Unaddressed participants' first gaze shift away from speaker A to speaker B, measured with respect to the end of the question (solid line = observed eye movements, dashed line = planned eye movements). The zero point on the *x*-axis (ms) marks the end of the question turn. The peak of the distribution represents the estimate of the mode. Dots represent the individual datapoints.

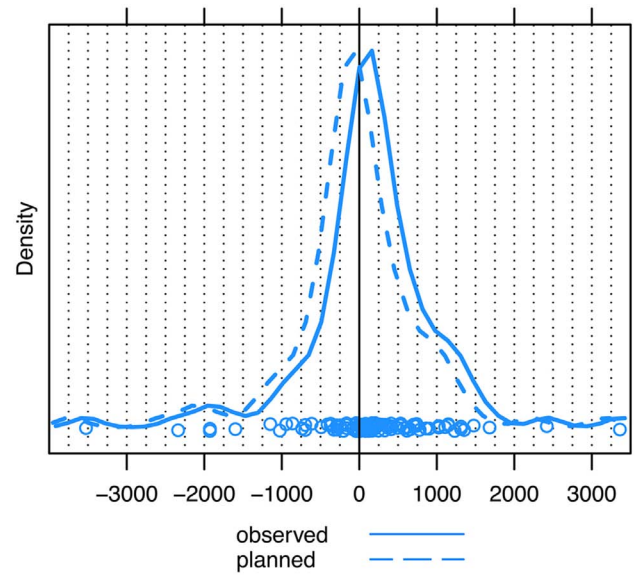


FIGURE 5 | Unaddressed participants' first gaze shift away from speaker A to speaker B, measured with respect to the first possible completion of the question (solid line = observed eye movements, dashed line = planned eye movements). The zero point on the *x*-axis (ms) marks the first possible completion of the turn. The peak of the distribution represents the estimate of the mode. Dots represent the individual datapoints.

first possible completion of each question; this corresponded to the end of the turn for those questions with only one possible completion. When plotting our data with respect to this reference point, the distribution yields a mode of about 160 ms just after the first possible completion (see **Figure 5**, solid line). Taking into account the 200 ms required to plan and launch observed eye movements, the distribution yields a mode of 40 ms just prior to the first possible completion (see **Figure 5**, dashed line).

When timing the gaze shifts with respect to the first possible completion of the question, we still see that a considerable number of gaze shifts from current to next speaker happen prior to the first possible completion, but much less so than when timing these gaze shifts with respect to the end of the turn: in 34.3% ($N = 36$) of cases, unaddressed participants shifted their gaze from current to next speaker before the first possible completion, and in 55.2% ($N = 58$) of cases unaddressed participants' gaze shifts had at least been planned prior to this point.

UNADDRESSED PARTICIPANTS' EYE MOVEMENTS AND ADDRESSED PARTICIPANTS' RESPONSES

Above we have shown that unaddressed participants are sensitive to first possible completions, as can be seen from the timing of their gaze shifts. However, considering that first possible completions mark points at which transition to the next speaker becomes relevant, addressed participants, too, are likely to be sensitive to these points and time their responses to them. This means that the first possible completion of speaker A's question and the onset of speaker B's response may often coincide. We therefore also measured the timing of the response[2] to see whether its onset may have attracted unaddressed recipients' attention and thus account for the timing of the gaze shifts we observed. And indeed, when we tested this statistically on our data, the result yielded a significant correlation between the unaddressed participants' first gaze shift from speaker A to speaker B and onset of speaker B's response [$\rho(13) = 0.234$, $p < 0.05$). This means that for responses that coincide with first possible completions of questions, gaze shifts could either be due to unaddressed participants recognizing the possible completion or reacting to the onset of the response. In order to tease these two factors apart, we carried out two further analyses by looking at two subsets of our data.

For the first analysis, we considered only those QR sequences where speaker A's first possible completion and speaker B's response onset did not coincide but where the response comes after the possible completion. For this comparison, we selected

[2]This measure considers vocalization as well as inbreaths produced in preparation for the response proper since preparatory inbreaths serve as pre-beginnings of turns (Schegloff, 1996) and thereby signal an intention to speak. Gaps and overlaps (i.e., gaps with negative values) were thus measured as the duration from the end of speaker's A turn to the beginning of speaker B's response, including audible inbreaths (see Materials and Methods). For sequences in which questions had only one possible completion (i.e., the first possible completion equated to the end of the turn), responses most frequently occurred after a gap of approximately 120 ms (Mode est. = 120 ms; Range = −943–1968 ms), with 31.3% (N=15) of responses in overlap with the question. For sequences in which questions had more than one possible completion (i.e., the first possible completion was not the end of the turn), the picture was predictably more complex, as responses could begin with reference to either of the two points. The distribution of gap durations in such sequences was non-unimodal, with a plateau between two slight peaks at −500 and 20 ms, and with 64.9% (N=37) of responses in overlap with the question.

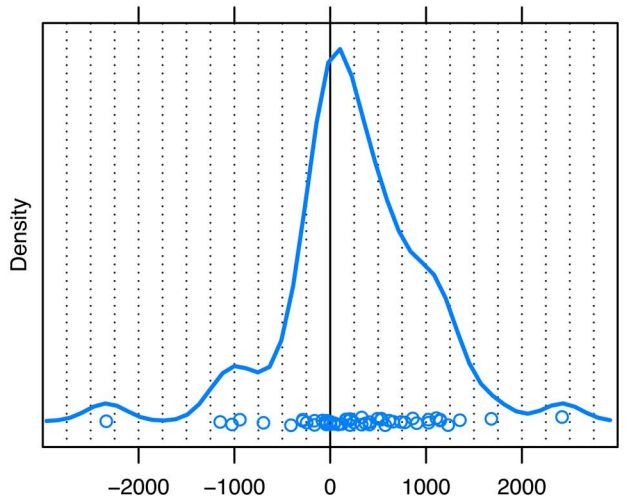


FIGURE 6 | Unaddressed participants' first gaze shift away from speaker A to speaker B for responses with an onset of 200 ms or more *after* the first possible completion of the question. The zero point on the *x*-axis (ms) marks the first possible completion of the turn. The peak of the distribution represents the estimate of the mode. Dots represent the individual datapoints.

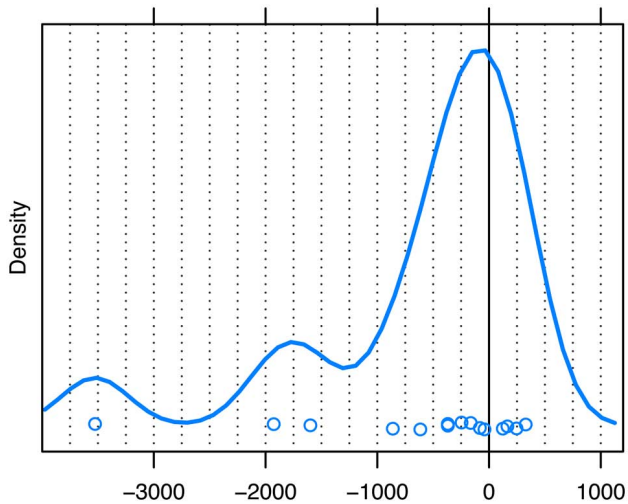


FIGURE 7 | Unaddressed participants' first gaze shift away from speaker A to speaker B for responses with an onset of 200 ms or more *prior* to the first possible completion of the question. The zero point on the *x*-axis (ms) marks the first possible completion of the turn. The peak of the distribution represents the estimate of the mode. Dots represent the individual datapoints.

those sequences where the response occurred more than 200 ms[3] after the first possible completion ($N = 54$). If the timing of the unaddressed participants' gaze shifts we observed based on the sample as a whole is explained by response onset rather than by first possible completions, then the mode for this subset of data should be at least 200 ms later than the mode for the distribution based on the entire sample. However, as can be gleaned from **Figure 6**, the mode for this subset is 105 ms (Range = –2337–2419), which differs only slightly from the mode of 160 ms for the entire sample. If anything, unaddressed participants' observed gaze shifts occur slightly earlier when B's response occurs 200 ms after the first possible completion, and certainly no later than when we consider the entire sample. Thus, unaddressed participants' eye movements in our data do indeed appear to reflect sensitivity to the first possible completion of the question, rather than being a mere reaction to the onset of the response.

However, we of course do acknowledge that response onset may also play a role in the timing of unaddressed participants' eye movements. In order to explore this further, we looked at another subset of our data, namely those cases in which speaker B's response began at least 200 ms before speaker A's first possible completion ($N = 15$). If response onset alone also attracts unaddressed participants' attention and, as a consequence, their gaze, then we should see that the mode of the distribution of gaze shifts for this subset is earlier than that for the distribution based on the sample as a whole. As can be seen from **Figure 7**, this was indeed the case, with the mode of observed gaze shifts for the subset of early responses being −35 ms, compared to an overall mode of 160 ms. This means that the eye movements within this subset must have been planned −235 ms before the first possible completion, which corresponds closely to the onset of these early responses at −200 ms or less.

[3]An earlier time point, such as 100 ms after the first possible completion, would also have been a viable comparison, but we settled for 200 ms since the hypothesized difference in modes would be clearer, and because the time needed to plan and execute an observable gaze shift would by that point have elapsed.

DISCUSSION

The present study sought to investigate the precise timing of unaddressed participants' eye movements during question–response sequences by advancing on previous research in two important ways. Firstly, we immersed the third person within the situated context of a spontaneous, live conversation in which he or she was a ratified participant (Goffman, 1979, 1981). Secondly, we aimed to zoom further into the pattern of anticipation established in previous studies by taking into account the intricate structure of questions in conversation. In addition, we here consider whether unaddressed participants' eye movements during question–response sequences are a turn-taking phenomenon *per se*, or whether they reflect processes of a different conversational order.

UNADDRESSED PARTICIPANTS TRACK CURRENT SPEAKERS

First of all, our analyses show that even when unaddressed participants are directly immersed in a conversation (rather than being third-person observers of pre-recorded dialogs), they move their gaze from one speaker to the next in about half of all question–response sequences. This confirms that, even when participating in actual 'on-line' interaction, unaddressed participants show a tendency to track current speakers (cf., von Hofsten et al., 2009; Foulsham et al., 2010; Tice and Henetz, 2011; Casillas and Frank, 2012, 2013; Edlund et al., 2012; Hirvenkari et al., 2013; Keitel et al., 2013), at least during QR sequences.

UNADDRESSED PARTICIPANTS SHIFT THEIR GAZE BEFORE TURN ENDS

When we examined the timing of gaze shifts with regard to turn ends, we found that the most frequent gaze shifts from current to next speaker were planned around 250 ms and observable around 50 ms prior to turn end. Thus, the results suggest that unaddressed participants' gaze shifts are predominantly anticipatory in nature when timed with respect to the end of question turns. As such, it replicates the findings from third-person perspective eye-tracking studies that have found evidence for adults shifting their gaze to the next speaker prior to turn end (e.g., Tice and Henetz, 2011; Casillas and Frank, 2012). Overall, this suggests that the third-person perspective eye-tracking paradigm, at least when used with adults and in the context of question–response sequences, serves as a good experimental proxy for investigating the eye movements of unaddressed participants in the context of turn-taking. A valuable follow-up to the present study may be one that employs the video recordings filmed from the unaddressed participants' view in a third-person perspective eye-tracking study as this would allow for a more direct comparison with the situated behavior to further corroborate this point.

However, despite the comparability, gaze shifts prior to turn end appear to be more common in actual conversation. In the present study, the majority of gaze shifts from current to next speaker occurred prior to turn end (60% of cases) or were planned and launched prior to this point (73% of cases). In contrast, in studies using the third-person perspective paradigm, either only a relatively small proportion of gaze shifts occurred prior to turn end (e.g., Tice and Henetz, 2011; Casillas and Frank, 2012) or none at all did (e.g., Edlund et al., 2012; Hirvenkari et al., 2013). Keitel et al. (2013) found that 54% of their adults' gaze shifts were anticipatory in nature – a proportion much closer to the present findings – but this percentage includes all gaze shifts made between 500 ms prior to the end of the current turn, as well as all gaze shifts made during the on average 900 ms gap preceding the next turn. It is therefore not possible to evaluate the extent to which these gaze shifts were anticipatory with respect to the end of the current turn, the measure we applied in the present study. In all, while gaze shifts do appear to be more anticipatory in actual conversation than in off-line eye-tracking paradigms, we have to consider that the present study focused exclusively on question–response sequences rather than on a mixture of different turn types. Since Casillas and Frank (2012) found a trend toward slightly stronger anticipation for questions than for non-questions for adults, we have to be mindful that this may also explain, or at least contribute, to the stronger pattern of anticipation found in the present study.

UNADDRESSED PARTICIPANTS SHIFT THEIR GAZE AT POSSIBLE TURN COMPLETIONS

The present study went further than just measuring eye movements with respect to turn ends. Here, we have taken into account the intricate structure of questions, and, more specifically, the first possible completion of each question, which for half of our questions was not the actual end of the turn. These points of possible completion create opportunities for a next speaker to take the turn, and it has been argued that participants in conversation are sensitive to these transition-relevance places (Sacks et al., 1974). Indeed, our data seem to corroborate this: we found that, in the majority of cases, unaddressed participants initiated the planning of their gaze shifts most frequently just 40 ms prior to the first possible completion of the turn. This time interval is shorter than the average duration of a single vowel in English (House, 1961; Umeda, 1975) and suggests that the planning of the most frequent gaze shifts more or less coincides with the point in the current turn at which transition between speakers first becomes relevant. Indeed, our measurement of the location of possible completions within a turn, which identifies them at the end of a word, is conservative. If the possible completion becomes recognizable even earlier, for example, as the result of an increase in the duration of final words or segments (see Local et al., 1986; Gravano and Hirschberg, 2011), the initiation of planning (i.e., the peak of the distribution in **Figure 5**) would occur after the possible completion, not before.

Thus, rather than a pattern of anticipation, in which unaddressed participants project the ends of question turns in advance, the virtual coincidence of possible completions and the onset of planning suggests that unaddressed participants recognize points of possible completion as they occur. That is, they seem to perceive specific cues closely associated with, and thus indicative of, the emergence of possible completions. Wells and Macfarlane (1998) have argued that transition relevance places can be defined in prosodic terms and that specific final major accents of a current turn signal its upcoming completion (cf. Schegloff, 1996, on 'pitch peaks' as indications of possible completion). They conclude that next speakers need not anticipate this accent; they merely have to recognize it. However, even the recognition of final accents or pitch peaks is a process that unfolds over time. The observation that gaze shifts are planned and launched 40 ms before the first possible completion of the current turn could therefore be interpreted as projection on a micro-scale, as it were, but it is something quite different from the long-range projection that has been argued for by some. Schegloff (1987) has proposed that the initial components of a turn can facilitate the projection of how it will end, *well before* it reaches possible completion (see also Levinson, 2013). This means that the grammatical structure of questions would allow unaddressed participants to shift their gaze to the next speaker at a very early point during the question. Considering that addressees are nonverbally responsive as speakers' utterances unfold (Clark and Krych, 2004; Bavelas and Gerwing, 2011; Traum et al., 2012), unaddressed participants may well feel inclined to gaze at the next speaker as early as possible to see how the emerging utterance is received. However, the present findings suggest that early projection of this kind does not govern the eye movements of unaddressed participants as they redirect their gaze from current to next speakers in question–response sequences. We do concede that unaddressed participants are likely to engage in sequence projection processes from very early on, which tells participants *what* is coming next (a response; Schegloff, 2007), and thus where to move their eyes (to the respective next speaker). However, local cues associated with the emergence of possible completion, rather than early turn projection, appear to act as a launch-signal by telling participants *when* to move their eyes.

Crucially, we have also shown that first possible completions govern unaddressed participants' gaze shifts in the absence of an early response. First possible completions alone appear to account for much of the data in our sample. (For 68.7% of our QR sequences the onset of the response came after the first possible completion.) At the same time, however, we have been able to show that early responses which precede the first possible completion also attract unaddressed participants' gaze, and thus may certainly be a contributing factor in those instances where first possible completion and onset of the response coincide. Hence, taking overlap between current and next speakers into account appears crucial if we aim to understand unaddressed participants' eye movements in natural conversation. Importantly, the effect of both factors – first possible completions and response onset – is based on a process of *recognition* rather than projection.

OPTIMIZING RECIPIENCY

To date, all studies of the phenomenon under investigation here – the redirection of gaze by unaddressed participants from current to next speakers at turn transitions – have used it to gain insight into processes involved in turn-taking. But the conclusion that the eye movements of unaddressed participants do not anticipate the first possible completion of the current turn, and thus do not necessarily reflect a projection of it, leads us to reconsider the nature of the phenomenon and to look elsewhere for principles that can account for the fine temporal coordination that we observe.

It has long been argued that among the many functions of gaze behavior in social interaction the use of gaze to display attention, engagement in the interaction, and recipiency to the current speaker is paramount (Goodwin, 1980, 1981; Heath, 1984, 1986; Kidwell, 1997; Robinson, 1998; Ruusuvuori, 2001; Ford and Stickle, 2012). Gazing at the current speaker not only shows one to be an attentive participant, whether directly addressed by the turn or not, but it also allows one to tap into the rich stream of visible behaviors that accompany turns at talk. Our results reveal that unaddressed participants redirect their gaze at a moment that is interactionally most optimal: by moving their eyes away from the current speaker not at the beginning of the question but close to its completion, unaddressed participants secure access to as much of the current speaker's visible bodily behavior as possible, including torso, head, and hand gestures, as well as lip movements and facial expressions that accompany the communicative action; at the same time, they also secure access to much of the next speaker's visual response to the question. Further, keeping their gaze on the current speaker until a very late point during the question allows unaddressed participants to display recipiency throughout most of the question, just as the reorientation to the addressed participant at the completion of the question allows them to do for the response. Both of these aspects, the visual behavior of speakers and its temporal coordination with possible turn completions, as well as the use of gaze for displaying and managing recipiency in multi-person interaction, are currently being investigated in more detail. This will help us to unravel the specific ways in which these factors contribute to the processing of turns and the organization of gaze in social interaction.

Although the gaze behavior of unaddressed participants does not necessarily reflect projection of the current turn, optimizing recipiency between current and next speakers does make use of the turn-taking system in other ways. Our results provide new and quantitative evidence that the recognition of points of possible completions are indeed core to the turn-taking system in conversation, as argued in Sacks et al.'s (1974) seminal paper. Moreover, it appears that not only addressed but also momentarily unaddressed participants orient to possible completions as they process turns at talk. This observation further underscores the point by Sacks et al. (1974, p. 727) that the organization of turn-taking creates an "intrinsic motivation for listening." One who wishes to have a turn at talk must attend to and process the current turn in order to recognize a point at which transition between speakers may occur. Even unaddressed participants, who do not take a turn in the question–response sequences in our data, show evidence in their gaze behavior of a fine attunement to this feature of the turn-taking system[4]. Our findings that unaddressed participants' gaze behavior during question–response sequences appears to be organized according to a principle that optimizes recipiency also fits well with the notion of an 'intrinsic motivation for participation,' as it were (Schilbach et al., 2010; Pfeiffer et al., 2014). Both Schilbach et al. (2010) and Pfeiffer et al. (2014) demonstrate that, in the context of gaze-based interactions, humans experience social-interactional engagement as rewarding, as evidenced by cerebral activity patterns in reward-related neurocircuitry.

In addition, it appears from our results that response onset can trump first possible completions, at least when these responses come prior to the first possible completion. In such cases, the timing of the response appears marked and may signal a marked social action (see Vatanen, 2014). That unaddressed participants orient their gaze toward the participant issuing a response of this status, despite the current turn not yet having reached its first possible completion, neatly fits the principle of optimizing recipiency.

The present study looked at eye movements with respect to one particular type of turn, that is, questions. Casillas and Frank (2012) found a marginally significant effect indicating that, in third-person perspective paradigms, adults show a stronger tendency to shift their gaze from current to next speaker – and a trend for this happening slightly earlier – for questions than non-questions. Corpus research on the timing of turn-taking in spontaneous conversation, however, found that participants responded as quickly to questions as to non-questions (Stivers et al., 2009). Further research on different types of turns is thus clearly needed and may help to explain why analyses that have combined questions with other turn types have not found evidence of anticipatory eye movements (Hirvenkari et al., 2013). Moreover, the present study focused on those question–response sequences that were associated with patterns of gaze behavior which would allow us to draw conclusions about unaddressed participants' cognitive processes relating to the anticipation of turn ends and upcoming responses. Question–response sequences associated with different

[4]Note that, while our analysis suggests a close coordination of eye gaze and turn structure, this does not exclude the possibility that participants' gaze is not also organized with respect to the structure of larger sequences and courses of actions (Rossano, 2012).

gaze patterns (such as unaddressed participants continuing to gaze at the questioner throughout the entire sequence) are not informative in this respect. Note that we are not suggesting that entirely different cognitive processes are at work in those cases. Quite the opposite – while it is very likely that unaddressed participants recognize possible turn completions also during those kinds of question–response sequences, other processes appear to be governing their eye movements causing them not to shift their gaze toward the responder at this point. What exactly these processes are is an open question and certainly worthy of future research, but they address a different question to the one under investigation here.

CONCLUSION

The present study has provided us with a first glimpse of the intricate connections between turns at talk and unaddressed participants' eye movements in spontaneous, multi-person interaction. On the one hand, we have here reproduced the basic findings from studies using the third-person perspective eye-tracking paradigm in spontaneous, live conversation. On the other, our data have provided us with stronger evidence that gaze shifts by unaddressed participants toward next speakers precede the end of the current turn than previous studies have. As such, our findings corroborate the notion that interactive paradigms do, at least in part, provide different insights than paradigms involving passive observation (Schilbach, 2010, 2014; Wilms et al., 2010; Pfeiffer et al., 2013; Schilbach et al., 2013). Further, the present study has advanced our understanding of which structures in the current turn guide unaddressed participants' eye movements in conversation and has helped to clarify the role that the projection of the current turn plays in this process. While our findings underline the general usefulness of third-person paradigms, they also point toward some of the limitations associated with this approach. Moreover, they point to the urgent need to consider not just actual turn ends but also first possible turn completions when measuring and interpreting eye movements during turns at talk. Finally, the present study has allowed us to identify a new interactional phenomenon, the optimization of recipiency, which appears to account for much of the gaze behavior of unaddressed participants during turn-taking.

ACKNOWLEDGMENTS

We would like to thank Linda Drijvers, Ludy Clissen, and Marloes van der Goot for their assistance with data preparation and coding, the Technical Group at the MPI for Psycholinguistics (especially Nick Wood, Ronald Fischer, and Reiner Dirksmeier) and SMI (especially Martin Pötter) for their technical advice and assistance throughout, Han Sloetjes (MPI) for his assistance with ELAN, as well as the participants who took part in our study. We would also like to thank the Language and Cognition Department (especially Stephen Levinson) and members of the Dialog Project at the MPI for their valuable feedback and discussion of this work, and the European Research Council (Advanced Grant #269484 INTERACT awarded to Stephen Levinson) and the Max Planck Gesellschaft for funding this research. We would also like to thank the two reviewers of our manuscript for their valuable input and thoughtful suggestions.

REFERENCES

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439. doi: 10.1006/jmla.1997.2558

Argyle, M., and Cook, M. (1976). *Gaze and Mutual* Gaze. Cambridge: Cambridge University Press.

Argyle, M., Ingham, R., Alkema, F., and McCallin, M. (1973). The different functions of gaze. *Semiotica* 7, 19–32. doi: 10.1515/semi.1973.7.1.19

Auer, P. (2005). Projection in interaction and projection in grammar. *Text-Interdiscip. J. Study Discourse* 25, 7–36.

Augusti, E. M., Melinder, A., and Gredebäck, G. (2010). Look who's talking: pre-verbal infants' perception of face-to-face and back-to-back social interactions. *Front. Psychol.* 1:161. doi: 10.3389/fpsyg.2010.00161

Bavelas, J. B., Coates, L., and Johnson, T. (2002). Listener responses as a collaborative process: the role of gaze. *J. Commun.* 52, 566–580. doi: 10.1111/j.1460-2466.2002.tb02562.x

Bavelas, J. B., and Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *Int. J. Listen.* 25, 178–198. doi: 10.1080/10904018.2010.508675

Becker, W. (1991). "Saccades," in *Vision and Visual Dysfunction*, Vol. 8, *Eye Movements*, ed. R. H. S. Carpenter (London: Macmillan), 95–137.

Boersma, P., and Weenink, D. (2014). *Praat: Doing Phonetics by Computer [Computer program]. Version 5.3.82.* Available at: http://www.praat.org/ [accessed July 26, 2014].

Bolden, G. B. (2013). Unpacking "self" repair and epistemics in conversation. *Soc. Psychol. Q.* 76, 314–342. doi: 10.1177/0190272513498398

Brône, G., and Oben, B. (2014). Insight interaction: a multimodal and multifocal dialogue corpus. *Lang. Resour. Eval.* doi: 10.1007/s10579-014-9283-2

Cary, M. S. (1978). The role of gaze in the initiation of conversation. *Soc. Psychol.* 41, 269–271. doi: 10.2307/3033565

Casillas, M., and Frank, M. C. (2012). "Cues to turn boundary prediction in adults and preschoolers," in *Proceedings of SemDial 2012 (SeineDial): The 16th Workshop on the Semantics and Pragmatics of Dialogue*, eds S. Brown-Schmidt, J. Ginzburg, and S. Larsson (Paris: Université Paris-Diderot), 61–69.

Casillas, M., and Frank, M. C. (2013). "The development of predictive processes in children's discourse understanding," in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Society), 299–304.

Clark, H. H., and Carlson, T. B. (1982). Hearers and speech acts. *Language* 58, 332–373. doi: 10.1353/lan.1982.0042

Clark, H. H., and Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *J. Mem. Lang.* 50, 62–81. doi: 10.1016/j.jml.2003.08.004

Cook, M. (1977). Gaze and mutual gaze in social encounters: how long—and when—we look others "in the eye" is one of the main signals in nonverbal communication. *Am. Sci.* 65, 328–333.

Couper-Kuhlen, E., and Ono, T. (2007). "Incrementing" in conversation. A comparison of practices in English, German and Japanese. *Pragmatics* 17, 513–552.

De Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

Duncan, S., Brunner, L. J., and Fiske, D. W. (1979). Strategy signals in face-to-face interaction. *J. Pers. Soc. Psychol.* 37, 301–313. doi: 10.1037/0022-3514.37.2.301

Edlund, J., Alexandersson, S., Beskow, J., Gustavsson, L., Heldner, M., Hjalmarsson, A., et al. (2012). "3rd party observer gaze as a continuous measure of dialogue flow," in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, Istanbul.

Fischer, B., and Ramsperger, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Exp. Brain Res.* 57, 191–195. doi: 10.1007/BF00231145

Ford, C. E., Fox, B. A., and Thompson, S. A. (1996). Practices in the construction of turns: the" TCU" revisited. *Pragmatics* 6, 427–454.

Ford, C. E., Fox, B. A., and Thompson, S. A. (2002). "Constituency and the grammar of turn increments," in *The Language of Turn and Sequence*, eds C. E. Ford, B. Fox, and S. A. Thompson (Oxford: Oxford University Press), 14–38.

Ford, C. E., and Stickle, T. (2012). Securing recipiency in workplace meetings: multimodal practices. *Discourse Stud.* 14, 11–30. doi: 10.1177/1461445611427213

Ford, C. E., and Thompson, S. A. (1996). "Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns,"

in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 134–184.

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., and Kingstone, A. (2010). Gaze allocation in a dynamic situation: effects of social status and speaking. *Cognition* 117, 319–331. doi: 10.1016/j.cognition.2010.09.003

Gardner, R. (2010). Question and answer sequences in Garrwa talk. *Aus. J. Linguist.* 30, 423–445. doi: 10.1080/07268602.2010.518554

Goffman, E. (1979). Footing. *Semiotica* 25, 1–30. doi: 10.1515/semi.1979.25.1-2.1

Goffman, E. (1981). "Footing," in *Forms of Talk*, ed. E. Goffman (Philadelphia, PA: University of Pennsylvania Press), 124–157.

Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociol. Inq.* 50, 272–302. doi: 10.1111/j.1475-682X.1980.tb00023.x

Goodwin, C. (1981). *Conversational Organization: Interaction Between Speakers and Hearers.* New York: Academic Press.

Gravano, A., and Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Comput. Speech Lang.* 25, 601–634. doi: 10.1016/j.csl.2010.10.003

Griffin, Z. M., and Bock, K. (2000). What the eyes say about speaking. *Psychol. Sci.* 11, 274–279. doi: 10.1111/1467-9280.00255

Heath, C. (1984). "Talk and recipiency: sequential organization in speech and body movement" in *Structures of Social Action: Studies in Conversation Analysis*, eds J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 247–265.

Heath, C. (1986). *Body Movement and Speech in Medical Interaction.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511628221

Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversation. *J. Phon.* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002

Hirvenkari, L., Ruusuvuori, J., Saarinen, V.-M., Kivioja, M., Peräkylä, A., and Hari, R. (2013). Influence of turn-taking in a two-person conversation on the gaze of a viewer. *PLoS ONE* 8:e71569. doi: 10.1371/journal.pone.0071569

House, A. S. (1961). On vowel duration in English. *J. Acoust. Soc. Am.* 33, 1174–1178. doi: 10.1121/1.1908941

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2:255. doi: 10.3389/fpsyg.2011.00255

Indefrey, P., and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001

Itier, R. J., and Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33, 843–863. doi: 10.1016/j.neubiorev.2009.02.004

Jefferson, G. (1986). Notes on "latency" in overlap onset. *Hum. Stud.* 9, 153–183. doi: 10.1007/BF00148125

Keitel, A., Prinz, W., Friederici, A. D., Hofsten, C. V., and Daum, M. M. (2013). Perception of conversations: the importance of semantics and intonation in children's development. *J. Exp. Child Psychol.* 116, 264–277. doi: 10.1016/j.jecp.2013.06.005

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol.* 26, 22–63. doi: 10.1016/0001-6918(67)90005-4

Kendon, A. (1990). *Conducting Interaction: Patterns of Behaviour in Focused Encounters.* Cambridge: Cambridge University Press.

Kendrick, K. H., and Torreira, F. (2014). The timing and construction of preference: a quantitative study. *Discourse Process.* doi: 10.1080/0163853X.2014.955997

Kidwell, M. (1997). Demonstrating recipiency: knowledge displays as a resource for the unaddressed participant. *Issues Appl. Linguist.* 8, 85–96.

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychol. Bull.* 100, 78–100. doi: 10.1037/0033-2909.100.1.78

Kobayashi, H., and Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *J. Hum. Evol.* 40, 419–435. doi: 10.1006/jhev.2001.0468

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310

Lerner, G. H. (2003). Selecting next speaker: the context-sensitive operation of a context-free organization. *Lang. Soc.* 32, 177–201. doi: 10.1017/S004740450332202X

Levinson, S. C. (1988). "Putting linguistics on a proper footing: explorations in Goffman's participation framework," in *Goffman: Exploring the Interaction Order*, eds P. Drew and A. Wootton (Oxford: Polity Press), 161–227.

Levinson, S. C. (2013). "Action formation and ascription," in *The Handbook of Conversation Analysis*, eds J. Sidnell and T. Stivers (Malden, MA: Wiley-Blackwell), 101–130.

Local, J., Kelly, J., and Wells, W. H. (1986). Towards a phonology of conversation: turn-taking in Tyneside English. *J. Linguist.* 22, 411–437. doi: 10.1017/S0022226700010859

Local, J., and Walker, G. (2012). How phonetic features project more talk. *J. Int. Phon. Assoc.* 42, 255–280. doi: 10.1017/S0025100312000187

Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., and Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *J. Cogn. Neurosci.* 26, 2530–2539. doi: 10.1162/jocn_a_00673

Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., et al. (2014). Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *Neuroimage* 101, 124–137. doi: 10.1016/j.neuroimage.2014.06.061

Pfeiffer, U. J., Vogeley, K., and Schilbach, L. (2013). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neurosci. Biobehav. Rev.* 37, 2516–2528. doi: 10.1016/j.neubiorev.2013.07.017

R Core Team. (2012). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna, Austria. Available at: http://www.R-project.org/

Robinson, J. D. (1998). Getting down to business talk, gaze, and body orientation during openings of doctor-patient consultations. *Hum. Commun. Res.* 25, 97–123. doi: 10.1111/j.1468-2958.1998.tb00438.x

Rossano, F. (2012). *Gaze Behavior in Face-to-Face Interaction.* Ph.D. thesis, Radboud University, Nijmegen.

Rossano, F., Brown, P., and Levinson, S. C. (2009). "Gaze, questioning, and culture," in *Conversation Analysis: Comparative Perspectives*, ed. J. Sidnell (Cambridge: Cambridge University Press), 197–249.

Ruusuvuori, J. (2001). Looking means listening: coordinating displays of engagement in doctor-patient interaction. *Soc. Sci. Med.* 52, 1093–1108. doi: 16/S0277-9536(00)00227-6

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.2307/412243

Salthouse, T. A., and Ellis, C. L. (1980). Determinants of eye-fixation duration. *Am. J. Psychol.* 93, 207–234. doi: 10.2307/1422228

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization With R.* New York, NY: Springer. doi: 10.1007/978-0-387-75969-2

Schegloff, E. A. (1987). "Recycled turn beginnings: a precise repair mechanism in conversation's turn-taking organisation," in *Talk and Social Organisation*, eds G. Button and J. Lee (Phiadelphia: Multilingual Matters), 70–93.

Schegloff, E. A. (1996). "Turn organization: one direction for inquiry into grammar and interaction," in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 52–133.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29, 1–63. doi: 10.1017/S0047404500001019

Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511791208

Schilbach, L. (2010). A second-person approach to other minds. *Nat. Rev. Neurosci.* 11, 449–449. doi: 10.1038/nrn2805-c1

Schilbach, L. (2014). On the relationship of online and offline social cognition. *Front. Hum. Neurosci.* 8:278. doi: 10.3389/fnhum.2014.00278

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., et al. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* 22, 2702–2715. doi: 10.1162/jocn.2009.21401

Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics* 6, 371–388.

Selting, M. (2000). The construction of units in conversational talk. *Lang. Soc.* 29, 477–517. doi: 10.1017/S0047404500004012

Senju, A., and Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends Cogn. Sci.* 13, 127–134. doi: 10.1016/j.tics.2008.11.009

Stivers, T., and Enfield, N. J. (2010). A coding scheme for question–response sequences in conversation. *J. Pragmat.* 42, 2620–2626. doi: 10.1016/j.pragma.2010.04.002

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587. doi: 10.1073/pnas.0903616106

Stivers, T., Enfield, N. J., and Levinson, S. C. (2010). Question–response sequences in conversation across ten languages: an introduction. *J. Pragmat.* 42, 2615–2619. doi: 10.1016/j.pragma.2010.04.001

Strömbergsson, S., Hjalmarsson, A., Edlund, J., and House, D. (2013). "Timing responses to questions in dialogue," in *Proceedings of the Interspeech*, Lyon, 2584–2588.

Tice, M., and Henetz, T. (2011). "Turn-boundary projection: looking ahead," in *Proceedings of the 33rd Annual Conference of the Cognitive Science Society,* eds L. Carlson, C. Hölscher, and T. Shipley (Austin, TX: Cognitive Science Society), 838–843.

Traum, D., DeVault, D., Lee, J., Wang, Z., and Marsella, S. (2012). "Incremental dialogue understanding and feedback for multiparty, multimodal conversation," in *Lecture Notes in Computer Science, Vol. 7502: Intelligent Virtual Agents,* eds Y. Nakano, M. Neff, A. Paiva, and M. Walker (Berlin Heidelberg: Springer), 275–288.

Umeda, N. (1975). Vowel duration in American English. *J. Acoust. Soc. Am.* 58, 434–445. doi: 10.1121/1.380688

Vatanen, A. (2014). *Responding in Overlap: Agency, Epistemicity and Social Action in Conversation.* Ph.D. thesis, University of Helsinki, Helsinki.

von Hofsten, C., Uhlig, H., Adell, M., and Kochukhova, O. (2009). How children with autism look at events. *Res. Autism Spectr. Disord.* 3, 556–569. doi: 10.1016/j.rasd.2008.12.003

Wells, B., and Macfarlane, S. (1998). Prosody as an interactional resource: turn-projection and overlap. *Lang. Speech* 41, 265–294. doi: 10.1177/002383099804100403

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., and Vogeley, K. (2010). It's in your eyes: using gaze feedback to create truly interactive paradigms for social cognitive and affective neuroscience. *Soc. Cogn. Affect. Neurosci.* 5, 98–107. doi: 10.1093/scan/nsq024

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN: A professional framework for multimodality research," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Nijmegen, 1556–1559.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 November 2014; accepted: 19 January 2015; published online: 09 February 2015.

*Citation: Holler J and Kendrick KH (2015) Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. Front. Psychol. **6**:98. doi: 10.3389/fpsyg.2015.00098*

This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology.

frontiers in Psychology

ORIGINAL RESEARCH
published: 26 August 2015
doi: 10.3389/fpsyg.2015.01113

Action-projection in Japanese conversation: topic particles *wa*, *mo*, and *tte* for triggering categorization activities

Hiroko Tanaka *

Department of Sociology, University of Essex, Colchester, UK

Conversation analytic work has revealed how anticipatory completions and preemptive actions can offer invaluable glimpses into the cognitive, contextual, grammatical, and temporal bases of projectability in turn-taking, by virtue of their potential not only as a display of participants' online prediction of roughly what it might take to complete a turn-in-progress but also to plan the next move. While the predicate-final word order and the incremental transformability of turns in Japanese generally lead to delayed projectability of turn-endings, this may be partially offset by the capacity of certain postpositional particles to trigger and propel prospective action trajectories. This article engages in a case study of the topic particle *wa* (and related particles *mo* and *tte*), by demonstrating how its grammatical affordances, the categorization activities, and cognitive processing it can set in motion, coupled with the immediate contextual, and temporal-productional features may coalesce to a point of critical mass, thereby enhancing the projectability of the not-yet-produced trajectory of the current turn. The discussion attempts to contribute to recent debates on ways language-specific lexicogrammatical resources are deeply interlinked with the types of opportunities that are provided for social action.

Keywords: conversation analysis, anticipatory completion, preemptive action, projectability, Japanese conversation, topic particle *wa*, membership categorization device, set theory

OPEN ACCESS

Edited by:
Kobin H. Kendrick, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Makoto Hayashi, University of Illinois, USA
Emi Morita, National University of Singapore, Singapore

***Correspondence:**
Hiroko Tanaka, Department of Sociology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ Essex, UK
htanaka@essex.ac.uk

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *16 February 2015*
Accepted: *20 July 2015*
Published: *26 August 2015*

Citation:
Tanaka H (2015) Action-projection in Japanese conversation: topic particles wa, mo, and tte for triggering categorization activities. Front. Psychol. 6:1113. doi: 10.3389/fpsyg.2015.01113

Introduction

The Phenomenon

The aim of this article is to demonstrate the potential for the situated use of the topic particle *wa* in Japanese conversation to serve as a powerful resource for locally projecting the possible trajectory of a turn-in-progress by activating and implementing a range of cognitive operations involving categorization activities. I focus mainly on the particle *wa* while touching upon related roles played by other particles including *tte* and *mo* (*wa* roughly glossed as "as for"; *tte* as "concerning"; whereas *mo* would be crudely equivalent to "also"). While the types of particles under consideration here are variously labeled "adverbial," "topic" or "focus" in the literature, they will be referred to as "topic particles" for simplicity (see Section Previous Research on *wa*). The abovementioned capacity of *wa* to strongly project action trajectories may be mobilized by participants in order to trigger and propel forward anticipatory completions and even preemptive actions through engaging in categorization activities, leading to a classification or re-classification of the universe of discourse.

As an illustration, in the following excerpt, the contingent use of *wa* (line 2) provides an opportunity for a coparticipant to implement a preemptive response (line 3). Four women have been asked to discuss their preferences in men. A participant L has characterized "narcissistic men" as "fun" to have as friends.

(1) [Sakura 07] Preferences

```
1    D:   tte  iu  ka tabun    atashi wa     kojinteki  ni:
          QUOT say QP probably I      as.for personally P
          'or, should ((I)) say, probably, as for me, personally,'

2         yo      de   iu  ikemen              wa- (.)
          society LOC  say good.looking.men   as.for
          'as for men who are conventionally considered good looking, (.)'

3 → F:    iya      da.
          dislike  COP
          '((I)) dislike ((them)).'

4    D:   chotto ne.
          little FP
          'a bit, you know.'
```

In line 1, D begins to formulate an assessment, by using *atashi* "me" and *kojinteki ni* "personally" to set the frame for the ensuing talk (Ono and Thompson, 2003, p. 332). She then introduces a referent "men who are conventionally considered good looking" (highlighted by a border) which is marked with *wa* (line 2). On hearing this turn-beginning (lines 1–2), F preemptively proffers her *own* assessment of the referent (line 3) by appropriating the grammatical slot made available by D's turn-beginning and constructing her turn as a grammatical continuation[1]. In other words, what D is projecting is being treated as so apparent that, for all practical purposes, it is seen to be sufficient not only for grasping D's intended action but for going one step further to formulate a response to it. D endorses F's action (line 4), thereby confirming F's understanding as implied in line 3. Through a close scrutiny of instances such as this where a *wa*-marked "reference formulation" (see Ford et al., 2013) triggers anticipatory completions or preemptive actions, I hope to shed light on the synergistic effect of the contextualization work performed by prior talk, the proximate temporal-productional features, and the grammatical and cognitive operations implemented by *wa*, for cumulatively laying the groundwork for augmenting the projectability of emerging turns.

The database for this study comprises approximately 20 h of telephone conversations and audio- or video-recorded face-to-face interaction among native speakers of Japanese, mainly from the Kanto or Kansai regions. Some of the data, including the Sakura corpus, are from publically available databases from TalkBank (MacWhinney, 2007), and relevant segments have been retranscribed by the author. Other data were collected by several different researchers in accordance with recommendations pertaining to human subjects of the local review boards of the universities to which they respectively belong. In each case, informed consent was freely given by all participants, and the data collected have been handled according to the Statement of Ethical Practice for the British Sociological Association (March 2002), including guidelines for the sharing of data collected for reuse in other projects. The excerpts selected for presentation in this article are drawn from the following conversations:

1. Sakura 07, Sakura 13, YKH 1, YKH 2 (video recordings of multi-party conversations)
2. IMD (telephone conversation)
3. Wedding Planning, MFriends (audio recordings of multi-party conversations)

Although space constraints limit consideration to nine excerpts, they are representative of recurrent patterns observed in the larger database. Please refer to the Supplementary Material for transcription notations and set-theoretic symbols used in this article. In the excerpts, boldface is used to highlight the topic particles under consideration, and the referents they mark are encased in a border.

Japanese Conversational Grammar and Projectability

This article builds on work in "interactional linguistics" (e.g., Ochs et al., 1996; Selting and Couper-Kuhlen, 2001; Thompson and Couper-Kuhlen, 2005; Couper-Kuhlen and Ono, 2007) and "projectability" in Japanese. Prior research has investigated the role of various grammatical elements for action projection in Japanese: a limited list including connectives (Mori, 1999), conjunctive particles (Hayashi, 1999; Lerner and Takagi, 1999; Tanaka, 1999), adverbials (Tanaka, 2001a), adverbial and case particles (Tanaka, 1999, 2005), complementizers (Maynard, 1993; Hayashi, 1997; Matsumoto, 1998; Tanaka, 2001b), final particles (Morita, 2005, 2012), postpositions (Hayashi, 2000, 2001, 2003, 2004), predicate-final structure (Nakamura, 2009), and micro-segmentation of units (Iwasaki, 2008, 2009, 2011, 2013a).

[1] F accomplishes this in part by leaving unexpressed any first person marking and partly through use of the copula *da* to mark the assessment as her own. Szczepek (2000a,b) discusses a similar practice in English of "borrowing" the construction of a prior turn to produce one's turn.

The above works show that projectability is closely connected with the structures of syntactic and prosodic resources of the language. Work in conversation analysis and allied perspectives in Japanese have shown that even though different word orders are preferred depending on the type of social action a turn is performing (e.g., Ono and Suzuki, 1992; Tanaka, 2005), there is nevertheless a predicate-final orientation in Japanese in the sense that the production of a predicate component is normatively treated as a possible transition-relevance place (Tanaka, 1999, 2000; Nakamura, 2009). Given that the action of a turn is often embodied within the predicate (Thompson and Couper-Kuhlen, 2005), the projectability of turns in Japanese is regularly delayed until the predicate has been produced (Fox et al., 1996; Tanaka, 1999). The limited projectability of turn-trajectories, however, is to some extent offset by the pervasive use of certain postpositional particles—"case" and "adverbial" particles in particular—which serve as resources for incrementally projecting the potential unfolding of a turn-in-progress (Tanaka, 1999). Case and adverbial particles are devices that retroactively specify the grammatical sense of the immediately preceding nominal that it "marks" (e.g., as a subject, topic, object, indirect object, etc.), and "establish a grammatical linkage with that nominal to form constituents of the form [nominal + postposition]" (Hayashi, 2004, p. 348).

Furthermore, case and adverbial particles (including topic particles) have the additional property of projecting some nominal or predicate component (a predicate in the case of topic particles) which may follow the particle within the local interactional environment (Tanaka, 1999), as schematized in **Figure 1**.

Applying **Figure 1** to excerpt (1), D's production of the referent *yo de iu ikemen* "men who are conventionally considered good looking" (line 2) together with the attachment of *wa*, forms a reference formulation NP + *wa* "as for men who are conventionally considered good looking," which serves a dual purpose here. First, the reference formulation provides a basis for F to predict a possible predicate that is being projected. Second, F proffers an agreement with what is predicted *Iya da.* "((I)) dislike ((them))," which is grammatically fitted as a continuation of the reference formulation.

Even though the marking of a nominal phrase with a topic particle thus opens up a grammatical slot for a forthcoming predicate, it is not always possible to project or predict with accuracy the kind of predicate that may be supplied (Tanaka, 1999, pp. 177–182). To wit, in spite of the ubiquity of *wa* within conversational interaction, most instances of *wa* do not in fact occasion anticipatory completions or preemptive actions. Despite the key role of grammar in turn-projection, it should be underscored that it is only one out of the range of resources coparticipants mobilize for predicting the possible turn-trajectory, most significantly the immediate interactional environment, sequential context, and productional features of the turn-in-progress (Lerner, 1991, 1996, 2004; Liddicoat, 2004).

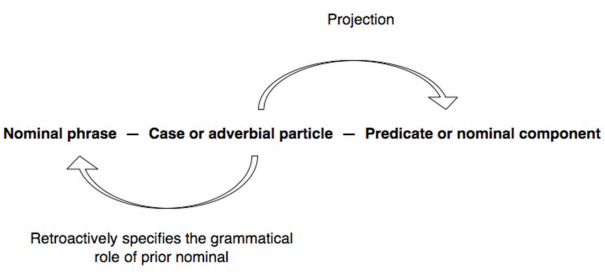


FIGURE 1 | Projective and retroactive properties of case and adverbial particles. (Adapted from Hayashi, 2004, p. 350; Tanaka, 1999, p. 155).

Studies on the social actions performed by postpositional particles have frequently touched upon the utility of *wa* for projecting the unfolding trajectory of turns (Hayashi, 1999, 2000, 2001, 2004; Tanaka, 1999, 2005). However, there is little in-depth research in conversation analysis specifically on the interactional role of *wa* or on the possibility that its projective capacity may hinge on the situated categorization activities it may be used to implement (but see Takagi, 2001).

Previous Research on *wa*

The potential roles and functions of *wa* in Japanese discourse have been discussed extensively by linguists, and have been at the center of countless debates, though a majority of the claims are based on invented or non-interactional data (see Shibatani, 1990, pp. 262–280; also Kuno, 1973; Clancy and Downing, 1987; Iwasaki, 1987, 2013b; Martin, 1987; Suzuki, 1995; Kaiser et al., 2001; Wlodarczyk, 2005, etc.). Shibatani (1990, p. 338) refers to *wa* and *mo* as "topic particles." Kaiser et al. classify *wa* and *mo* as "focus particles" but distinguish the two by suggesting that whereas *mo* focuses on the nominal that it follows, *wa* primarily focuses on the predicate that follows (Kaiser et al., 2001, p. 577). They add, "*wa* is often called a topic P (particle), because it typically marks the topic of a topic-comment type S (sentence). The focus in these S again is on the comment or pred(icate)." (Kaiser et al., 2001, p. 577, parentheses added). In relation to broader grammatical groupings, Tsujimura (1996, p. 134) sees the topic particle *wa* as a type of case particle, but Shibatani (1990) distinguishes case from adverbial particles, and classifies *wa*, *mo,* and *tte* as adverbial particles. The particle *tte* is variously called a "quotative particle," a "definition particle" (Kaiser et al., 2001) or a "complementizer" (Matsumoto, 1998). Depending on the particular usage, it has been described as being equivalent to other forms such as *to*, *to iu*, *to iu no*, or *to iu no wa* (see Kaiser et al., 2001).

In a well-known work, Kuno (1973, pp. 44–49) posits two types of *wa*: the "thematic" and "contrastive" *wa*. This position is contested by Shibatani, who argues that both functions can be subsumed under the rubric of the contrastive *wa*, but that the contrast "only becomes apparent when a parallel or contrasting proposition exists overtly or covertly" within the discourse environment (Shibatani, 1990, p. 265). Others like Martin (1987, pp. 60–65) and Kaiser et al. (2001) enumerate multiple usages for the particle, while noting that one such usage is to mark contrasts. For instance, according to Kaiser et al. (2001, p. 582), when a comment is made on a nominal (phrase) marked with *wa*, it "implies that the comment may not apply to other" nominal

(phrases). This raises the issue of specifying the kinds of "other nominals" that the comment would be inapplicable to. Another frequently reported feature of *wa* is that its usage and that of the particle *mo* are "mutually exclusive" (Takeuchi, 1999, p. 133). While *wa* is purported to have the general characteristic of "excluding" the nominal phrase that it marks, *mo* is described as "inclusive" and is translated as "too" or "also" (Kaiser et al., 2001, p. 242).

Maruyama (2003) addresses some of the issues indicated above by examining the function of *wa* in naturally occurring conversation, focusing on the importance of the discourse context in which *wa* occurs. She reports that a majority of cases of *wa* in her data fall into two main types of schemata, both of which mark a contrast (an opposite or parallel relationship) in some way: in the first type, given a component Y which contains a *wa*-marked nominal, attention on the discourse context prior to Y yields components X which stand in a semantically contrastive relation to Y; as for the second type, likewise given a component Y containing a *wa*-marked nominal, the discourse context prior to Y will contain a Set X comprising various components from which the component Y is specifically being singled out. With respect to the latter type, she notes that although X and Y do not stand in semantically contrastive relation, "*wa* in this schema still marks a contrast, for when Y is chosen out of the Set X by a speaker, Set X and Y are in a contrastive relationship in the sense that only Y is chosen" (Maruyama, 2003, p. 268). It is becoming common in recent commentaries on *wa* to incorporate the concept of "sets" within the explanatory apparatus, as exemplified by Shoichi Iwasaki's characterization of the contrastive function of *wa* to mark a referent to "represent an entity that is set off against another entity of the same class…due to their different attributes, which nonetheless constitute a coherent set" (Iwasaki, 2013b, p. 244).

Research in interactional linguistics is increasingly converging on the notion that postpositional particles primarily have a pragmatic rather than a grammatical role (e.g., Ono et al., 2000 on *ga*). Following in this vein, Takagi's (2001) study of child-adult interaction focuses on the use of *wa* in question formulations of the form "referent + *wa*?" (which she refers to as "*wa*-ending turns"). Takagi argues that a *wa*-ending turn is simultaneously deeply embedded in the particulars of the ongoing activity while at the same time prospectively oriented by inviting a recipient to supply a predicate that will be associated with the referent marked by *wa* (Takagi, 2001, p. 187). What is more, she contends that a *wa*-ending turn invariably has a directionality (not observed with other particles such as the nominative *ga* or accusative *o*) which propels the sequence forward by providing a grammatical slot for recipients to offer "what can be said about the reference," and going beyond simply "projecting" what should come next (Takagi, 2001, p. 187). Drawing on this and other previous studies, the present article pays particular attention to the contingent treatment of *wa* as mutually displayed by participants within the dynamic moment-by-moment unfolding of talk. In doing so, insights may be gained into its extensive utility for (membership) categorization activities. It will be shown that *wa* and other topic particles are critical resources for the performance of rudimentary categorization operations.

Membership Categorization and Set Theory

Membership categorization (Sacks, 1972, 1986) is concerned with practices used by participants in interaction to categorize people and the activities they engage in. In the process, participants display their cultural knowledge and commonsense reasoning in understanding and classifying the social world around them. Sacks points out that there are various membership categories that are used in our everyday interaction—such as the set of members of a population who are professionals. Moreover, there are certain ways in which we associate particular categories with others because they "go together" in some way—e.g., the larger class consisting of two categories, professionals and laypersons, which we associate together because they classify persons according to whether they have special rights to deal with certain types of troubles or not. Sacks calls such overarching classes "membership categorization devices" or MCDs:

> By this term I shall intend: any collection of membership categories, containing at least a category, which may be applied to some population containing at least a member, so as to provide, by the use of some rules of application, for the pairing of at least a population member and a categorization device member. A device is then a collection plus rules of application (Sacks, 1986, p. 332).

An often cited example of a MCD is one defined along the dimension of "stages of life." If we denote membership categories by using curly brackets { } and a membership categorization device through square brackets [ ], the MCD "stages of life" consisting of different membership categories may be represented by [{babies}, {toddlers}, {children}, {adolescents}, {young adults}, {the middle-aged}, {the elderly}] or through a relative measure in relation to the ego as in [{younger persons}{older persons}]. It should be noted that these collections are not analytical categories, but are invoked by participants to reflect members' knowledge as contingently formulated and locally negotiated in interaction.

As will be discussed herein, *wa* (as well as other topic particles including *tte* and *mo*) are implicated in the performance of the most primordial of membership categorization or set-theoretic operations (see Wlodarczyk, 2005). The data reveal that topic particles are employed to classify all manner of things in the physical and conceptual universe. Indeed, it has been suggested that these resources are used "indiscriminately" whether they apply to person, object or conceptual categories.

> While people certainly differ from objects as stimuli, the categorization rules and conceptual structures used in person and object perception may not be fundamentally different. Moreover, to the degree that differences do exist we can, presumably, gain finer insight into person categorization systems by comparing and contrasting them against this baseline of object categorization (Cantor and Mischel, 1979, p. 8).

In order to make full use of prior research on membership categorization in conversation analysis while simultaneously drawing on notions from rudimentary set theory (e.g., Halmos, 1960), only excerpts bearing on person references and categories will be used as examples in this article, though it can be empirically established that much of membership categorization is extendable and adaptable to other types of categories and collections of categories. Thus, the term "category" will be used interchangeably with "set," and "membership categorization device" as equivalent to the notion of the larger collection that contains the categories/sets which are associated together along some dimension.

In the half century following the inception of conversation analysis, the insights provided in Sacks' (1972, 1986) seminal work on membership categorization have been further developed by conversation analysts and ethnomethodologists (Hester and Eglin, 1997; Egbert, 2004; Schegloff, 2007a,b; Deppermann, 2011; Lerner et al., 2012; to name but just a few). The reader is referred to Day (2013) for a useful summary. The journal special issue [*Discourse Studies* 2012 Issue 14(3)] is a reflection of a renewed recent interest in membership categories.

The following sections proceed step-by-step to construct a picture of the ways in which members use *wa* (and other topic particles) for performing categorization or set theoretic operations and projecting the upcoming trajectory of talk. A range of interactional environments in which the situated marking of a referent with *wa* triggers anticipatory completions or preemptive actions will be examined, suggesting a close interconnection between the kinds of categorization work that *wa* can perform, the nature of the prior contextualization work, and the temporal-productional features of talk.

Basic Categorization/Set-theoretic Actions Performed by Topic Particles

Before narrowing the focus to the role of *wa*, it would be useful to gain a sense for how members may deploy a range of topic particles as interactional resources depending on the kind of categorization activity to be implemented. I begin with *tte* as a typical example of a topic particle that can contribute to laying the groundwork for further categorization activities, and go on to discuss the mutually exclusive uses of *mo* and *wa*. The particle *tte* shares with other topic particles the general characteristic of marking a referent and projecting a predicate. It will be shown that one of the relevant activities *tte* may engender is to topicalize the incumbency of a referent in some category.

The way *tte* operates on the parameters "referent" and "category" is illustrated in the following excerpt taken from a telephone conversation between fellow alumni from high school, Ken and Mai. Ken has called Mai to tell her about a grandiose wedding reception he attended recently in which Yoko, a common friend of the two from high school, was the bride. From an earlier part of the conversation, it is clear that Yoko is a medical doctor, and that she is marrying another doctor from the same university hospital. Immediately before the part shown, Ken has been describing the guests attending the reception. There is something in Ken's telling which Mai notices as departing from her presupposition, as indicated by her turn-initial *eh!* in line 1 (see Hayashi, 2009).

(2) [IMD 254] Doctors

```
1   Mai:   eh! .hh ano Yoko (.)  tte        naika             deshoo:?
           what    uhm ((name))  concerning internal.medicine COP
          'what! .hh uhm concerning Yoko, ((she)) is ((in)) internal medicine,
           isn't ((she))?'

2   Ken:   'n::
          'mm:'

3 →Mai:    dan'na san wa:?=
           husband    as.for
          'what about ((her)) husband?'

4 →Ken:   =mmo naika             tte  yutteta     yo.
          also internal.medicine QUOT were.saying FP
         'is also ((in)) internal medicine, ((they)) were saying'

5   Mai:  a::u- ah! on'naji naika              na[ n ka:.=↑aa:]::=
          oh::  oh! same    internal.medicine  COP N QP    oh
          oh:: oh! so ((he))'s ((in)) the same internal medicine.=oh:::=
6   Ken:                                        [  'n:::::.  ]
                                                   'mm::::.'

7   Mai: =↑dakara↓ ka:.=hee:::::.
           why     QP   wow
        ='that's why=wo:::::w.'
```

Mai's question in line 1 is tantamount to asking for confirmation that Yoko is an incumbent of the category {doctors of internal medicine}. The capacity of *tte* to invoke the relevance of membership in a category draws in part on "the economy rule" that "if a member uses a single category from any membership

categorization device, then they can be recognized to be doing *adequate reference* to a person" (Sacks, 1986, p. 333).

The use of *tte* for assigning membership of a referent in a category concomitantly proposes "classifying things" as a relevant activity to be engaging in within the local context, as demonstrated by the regularity with which such instances either engender, or are used as a preliminary to, some main categorization activity. Once the groundwork is established, co-participants can exploit it as a framework to engage in further categorization activities, by activating "the consistency rule": "If some population of persons is being categorized, and if a category from some device's collection has been used to categorize a first member of the population, then…other categories of the same collection *may* be used to categorize further members of the population" (Sacks, 1986, p. 333). In the present case, line 1 sets the stage for classifying another member, as instantiated by Mai's main query (line 3) to be examined closely below. Although it is not possible to elaborate here, other topic particles such as *toka* invoked (i.e., internal medicine) is partly indicated by the fact that Mai does not use the equally accessible alternative question formulation *dan'na san mo?* "The husband also?" [see Excerpt (3) line 5 for an example], as well as by the way Mai subsequently responds to Ken's answer to the question. Further evidence of the potential tilting toward a different category of this situated use of *wa* will be examined below.

But first, we see that Ken goes on to respond that the husband is *also* in internal medicine (line 4), by countering Mai's presupposition. In order to do this, he has been forced to adopt a turn-beginning that avoids the particle *wa*, which can be used to project possible "exclusion" of the husband from the category {doctors of internal medicine} (line 3), and instead, to use *mo* which projects "inclusion" in the same category (line 4). By beginning with *mo*, Ken constructs a "postposition-initiated utterance" (Hayashi, 2000, p. 215ff) which connects with the same referent *dan'na san* in Mai's query (line 3) and now marks it with *mo* (line 4), thereby altering the trajectory of the turn:

```
3   Mai:   dan'na san wa:?=
           husband   as.for
           'what about ((her)) husband?'

4   Ken:   =mmo      naika              tte  yutteta     yo.
           also      internal.medicine  QUOT were.saying FP
           'is also ((in)) internal medicine, ((they)) were saying'
```

The procedure results in marking *dan'na san* not with *wa* but *mo*:

dan'na san **wa** ----------▶ *dan'na san* **mo**

(see excerpt (7) line 4) and even *wa* (see excerpt (8), line 6) may likewise be used for proposing categorization as an activity to be pursued.

Crucially, this excerpt also illustrates the mutual incompatibility of the operations performed by *wa* and *mo* respectively, at least in the specific context where a category (in this case, {doctors of internal medicine}) has just been invoked. Specifically, after receiving the sought-after confirmation that Yoko specializes in internal medicine (line 2), Mai next proceeds to ask about Yoko's husband, *dan'na san wa:* "what about ((her)) husband?" (line 3), by using a question formulation that exploits the projective properties of *wa*. As noted in Section Previous Research on *wa*, this use of *wa* serves as "an invitation to provide what can be said about the reference in the *wa*-ending turn" (Takagi, 2001, p. 187). First, lines 1 and 3 taken together propose that Mai knows that the husband is likewise a medical doctor but not his specialty, since it is the specialty that is the target of the query. Furthermore, it can be argued that the employment of the *wa*-ending turn, *dan'na san wa:* in this specific sequential context, namely, immediately following the invocation of a category, exhibits Mai's presupposition that the husband is more likely than not to have a different specialty from that of Yoko—i.e., that the husband is potentially a member of a category {doctors of specialty *Y*} where *Y* is unspecified but different from internal medicine. The possible tilting toward the husband belonging to a different category than the one already now enabling Ken to project with "consistency" that the husband's specialty is the same as Yoko's. That Mai may have not even contemplated such a "coincidence" (when she initiated her enquiry through the use of *wa* in line 3) is displayed in her uptake in lines 5 and 7: through the repeated deployment of *aa* "oh" to index a "change of state" (Heritage, 1984) from not knowing to being informed; through commentary attributing the "change of state" to the revelation that the husband is likewise in internal medicine; and finally, through the interjection *hee* "wow," proposing that Ken's informing has resolved some incongruity that had been puzzling her in line 1 (see Tanaka, 2013). These observations reinforce the possibility that a question formulation *x wa?* immediately following an invocation of a category *Y* may contingently be tilted toward an answer that excludes *x* from the category *Y*, although further work is needed to explore its workings in other local contexts.

Thus, in terms of categorization activities, lines 3 and 4 exemplify three basic operations performed by *wa* and *mo*. First, by deploying *wa* to mark the referent *dan'na san* "husband," Mai potentially excludes the referent from the already invoked category {doctors of internal medicine} *and* suggests that the husband may belong to a different, though unspecified category {doctors of specialty *Y*} which Ken is invited to name. Second, Mai's deployment of *wa* additionally invokes an overarching membership categorization device "types of medical doctors" in which the respective categories to which Yoko and her husband

may belong to are co-class members, through an application of the "consistency rule." Third, whereas Mai's turn (line 3) potentially places the husband outside the category {doctors of internal medicine}, Ken returns the husband in the category {doctors of internal medicine}. The entire process is schematized in **Figure 2**.

As a further demonstration of the differential usages of *wa* and *mo* and the possible tilting of a *wa*-ending question formulation toward a category different from one that has been invoked, I reanalyze Excerpt (3) from Takagi (2001), which shows a very young child Y (2 years and 4 months) switching between *wa* and *mo* to index her evolving expectations as to who (among the people present in the room) may be participating in a planned visit to her grandparents' house in a few days time—i.e., inclusion or exclusion from the category {people who are going on the visit}. The little girl is asking her mother M as to who will be going on the outing. *Jun-kun* is her brother (5 years and 2 months). In line 10, the child is referring to the researcher (a stranger) who is visiting for the purpose of making recordings of the family interaction. Note that *–kun* and *–chan* are informal name suffixes commonly used when addressing or referring to someone (or oneself in the case of a small child).

Using *wa*-ending question formulations, Y begins by asking whether her father, mother and brother are going on the visit. Once it is established that all other members of her immediate family will participate (i.e., members of the category {those going on the visit}), Y then switches to the use of *mo* (line 5) to enquire about herself, displaying a "reasonable" assumption of the likelihood of herself being included in the said category. In the absence of an immediate affirmation (line 6) however, Y "repairs" her *mo*-ending question formulation, with its tilting toward "inclusion," to the *wa*-ending (line 7), which divests the question of such an assumption and is tilted instead toward the co-class category {those who are not going on the visit}. It nevertheless emerges (line 8) that Y was justified after all in assuming inclusion in the former category (line 8). Interestingly, Y avoids using *mo* when next enquiring about the researcher (line 10), thereby exhibiting her assumption that the researcher is unlikely to participate in the family visit. The child appears to be using *mo* and *wa* to display her differential predictions (and deductive processes) with regard to probable inclusion or exclusion: *mo* to index an expectation for a referent to be included in the previously invoked category, and *wa* for the converse (i.e., inclusion in the co-class category).

(3) Yacchan (from Takagi, 2001, pp. 158–159; modified translation)

1 Y: *papa* ***wa***:: *mama* ***wa***::?
Daddy as.for Mommy as.for
'is Daddy ((going))? is Mommy ((going))?'

2 M: *papa* ***mo*** *mama* ***mo*** *iku yo*:
daddy too mommy too go FP
'both Daddy and Mommy are going.'

3 Y: *Jun kun* ***wa***?
((name)) SFX as.for
'what about Jun-kun?'

4 M: *Jun kun* ***mo*** *iku yo*:
((name)) SFX also go FP
'Jun-kun is also going.'

5 → Y: *Ya cchan* ***mo***? ((asking about herself))
((name)) SFX also
'Yacchan too?'

6 M: *mhhhhhhhh*

7 → Y: *Ya cchan* ***wa***?
((name)) SFX as.for
'how about Yacchan?'

8 M: *Ya cchan* ***mo*** *iku yo*.
((name)) SFX also go FP
'Yacchan is also going.'

9 (1.5)

10 → Y: *kore*: *kono hito* ***wa***? ((referring to the researcher))
this this person as.for
'how about this this person?'

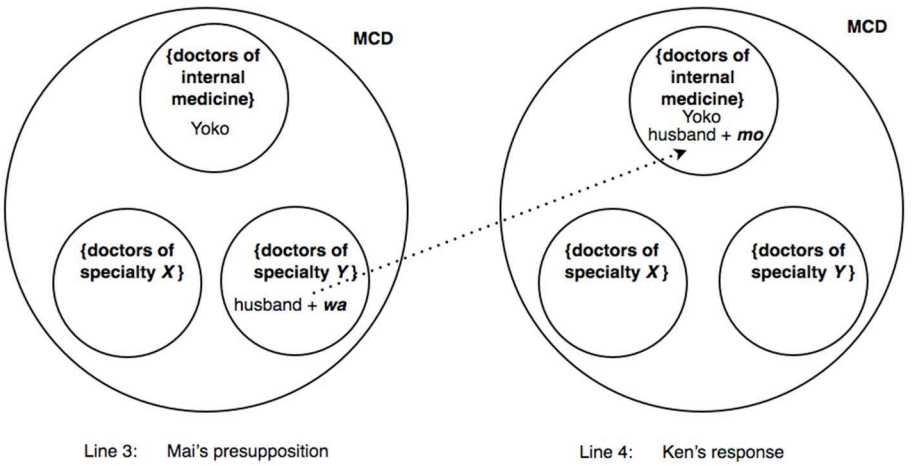


FIGURE 2 | Excerpt (2): In enquiring about the husband's specialty, Mai marks "the husband" with *wa*, which has a tilting toward exclusion from {doctors of internal medicine} and toward inclusion in {doctors of specialty *Y*}, where *Y* is unspecified (left). Ken responds by marking "the husband" with *mo* to place him back in {doctors of internal medicine} **(right)**. MCD = "types of medical doctors" = [{doctors of internal medicine}, {doctors of specialty *Y*}, {doctors of specialty *X*},...], where *Y* and *X* are unspecified. Only three of the many possible specialties have been represented in the figure.

Another excerpt is considered to provide a recipient's perspective on the possibility that a question formulation *x wa?* (immediately following the invocation of some category or formulation *x wa?* Four male university students have been asked to talk about their preferences in women. The discussion has digressed from desirable character traits to physical attributes:

(4) [Sakura 13] Kindness

```
1    K:    ashiha     to  muneha      da  na.
           legs.camp  and bosoms.camp COP FP
           'it'll be the legs camp ((for H)) and the bosoms camp ((for me)).'

2    G:    jaa  ore shiriha   de.
           then I   bums.camp INS
           'then I'll ((go)) for the bums camp.'

3    K:    J wa?
           J as.for
           'J, what about ((you))?'

4 →  J:    nani  nokotton  no, [ato. heh heh heh heh
           what  remaining QP   else
           'what else is there left?'
5    All:                      [((laughter))

6    K:    yasashisa de.
           kindness  INS
           '((how about going)) for kindness.'

7    J:   [yasashisa de. ((nod)) ...
           kindness  INS
           '((I'll go)) for kindness.'
8    All: [((laughter))
```

categories) may contain an implicit tilting toward the category incumbency of *x* in a different co-class category of the ones already invoked. In excerpt (4), a participant makes explicit his interpretation of the categorization implications of a question

K's turn in line 1, in effect, assigns H and himself respectively to the categories "legs camp" and "bosoms camp," in the MCD of men categorized according to their (anatomical) preferences. G follows suit in line 2, putting himself in the "bums camp,"

using the connective *jaa* "then" to indicate that he has limited his choice to a not-yet-selected camp. K then turns to J through the question formulation *J wa?* (line 3). Interestingly, J responds with a playful counter-question "what else is there left?" (line 4), thereby exhibiting an interpretation of the question formulation *J wa?* as embodying an implicit expectation to select a camp (category) not claimed by the others—namely, a co-class category in the same MCD.

In (5), *wa* is mobilized in a similar interactional environment, but in this case, for implementing and confirming an understanding check. Importantly, the excerpt exemplifies the interlocutors concurring (through co-constructions) on the action potentially being projected by *wa* in the immediate aftermath of the invocation of a category, thereby making evident an implicit tilting of the employment of *wa* toward exclusion of the referent it marks from an already invoked category. W and her fiancée H are arranging the logistics of their wedding reception, guided by S, their wedding planner. S has told W and H that it is more customary to provide a single take-home gift for guests who are a married couple rather than separate gifts. To this, W has just mentioned that she knows of cases whereby wives receive alternative gifts. In line 1, she is asking about such gifts.

(5) Wedding planning

```
1     W:    sooyuu      no w- wa
            those.kinds GEN   TOP
           '((do you mean that)) as for those kinds of things ((i.e., the
            alternative gifts)),'

2           zenbu:     danna san no  hoo  ni  tsuke[te:
            everything husband   GEN side LOC attach.and
           'everything would go to the husband, and'
3     S:                                          [tsukete:
                                                   attach.and
                                                  'would go to, and'

4     W:    okusama wa:[
            wife    as.for
           'as for the wife'
5 →   S:                 [wa? ((glottal stop))
                          as.for
                         'as for'

6 →   W:    nashi  [toka.
            nothing e.g.
           '((for her to get)) nothing, for instance'
7 →   S:           [tokuni       nashi   tte  yuu no ga
                   in.particular nothing QUOT say N  SUB
                   '((for her to get)) nothing in particular,'

8 →         ooi   desu keredomo:
            often COP  CONJ
           'is more common.'
```

In lines 1–2, W first checks if S is implying that such alternative gifts should likewise all go to the husband—i.e., that the husband belongs to the category {guests receiving all the take-home gifts}. S affirms this through a co-construction (line 3). Then, in lines 4 and 6, deploying the reference formulation *okusama wa:*, W embarks on a further understanding check as to whether the wife tends to get nothing—namely, that the wife may belong to the co-class category {guests receiving no take-home gifts}. In lines 5 and 7–8, S affirms W's understanding again by co-constructing W's turn. A closer inspection of the intricate, moment-by-moment coordination of action here affords a rare opportunity to witness the action-projection-capacity of *wa* being ratified and jointly mobilized for implementing the categorization activity of exclusion from an already invoked category. First, on hearing W's talk *okusama wa:* (line 4), S quickly echoes simply the *wa* (line 5), thereby endorsing and herself re-mobilizing its capacity for projecting the trajectory of the ongoing turn. W, for her part, treats S's echoing of *wa* as a go-ahead to render explicit what *wa* is being used to project (line 6), duly ratified by S (lines 7–8). In other words, the speakers are collaboratively displaying and implementing their shared understanding of the use of *wa* for excluding a referent from the category which was invoked immediately beforehand.

This section has demonstrated members' orientations to *tte* as a resource to invoke a category, and the mutual exclusivity of *wa* and *mo* depending on the type of categorization activity being proposed. In brief, *wa* is contingently used to exclude a referent from a previously invoked category (which thereby makes relevant a different category in a MCD), and to assign it

to a co-class category within that MCD. On the other hand, *mo* is used to mark a referent and to include it in a category which has already been invoked.

Wa for Triggering Anticipatory Completions and Preemptive Actions

Observations were made above concerning the types of categorization activities that may be performed through a number of topic particles. Among other things, it was shown that the use of *tte* for explicit invocation of a category and a MCD is one way of providing a foundation, which participants may build on to perform further categorization work such as exclusion or inclusion of other members of the population from the said category. Needless to say, employing *tte* is not the sole way to realize such prior contextualizing work. The instances to be considered show that critical groundwork may be laid in a variety of other ways through participants' coordinated mobilization of resources that emerge contingently within the unfolding of talk. The aim here is to explore how such preliminary activities can give rise to an interactional environment ripe for the situated deployment of *wa* that activates anticipatory completions and preemptive actions.

In contrast to the verbally explicit invocation of a category, for instance in excerpt (2) above, excerpt (6) below exemplifies how precursory categorization work may be initiated through visual conduct even when no category is named, and spark off further categorization activities. Moreover, it provides additional empirical support for the mutually exclusive operations of *wa* and *mo*. Recall that in excerpt (2), *mo* was used in order to include a referent in a category that had already been invoked, to reverse an apparent presupposition about category non-incumbency suggested by *wa*. The converse is demonstrated here: namely, how *mo* may be replaced by *wa* in order to repair a presumption about category incumbency displayed through *mo*.

Chie has been engaging in a telling about a recent holiday at a Hawaiian theme park, which Mari has no knowledge of. Shortly before the extract, Chie has begun to describe the Hawaiian shows that were featured. In the part shown, Mari is prompting Chie to elaborate. Included in the transcript are descriptions of some visual conduct critical for understanding the categorization activities the participants implement.

(6) [YKH 2, 3'43-4'02"] Hawaiian show

```
1  Mari: ja   utsukushii on'na no hito |ga,
         then beautiful  women          SUB
        'so were there beautiful women who
         ((marked as grammatical subject))'
                                      |((lifts up hands))

2  Chie: so↓ o rai[toappu |sareteru shi:, (.)
         yeah  light.up    PASS     CONJ
        'yeah, were illuminated by
         spotlights and (.)'
                         |((lifts up arms))
3  Mari:          [koo,
                  'like this'

4  Chie:  |koo      [yatteru wake yo.
           like.this doing   FP   FP
          'going like this.'
          |((enacting arm motions of hula dance))
5  Mari:            [°°he:o::°°
                    'wo::w'

6  Mari:   <de  otoko no hito  mo:=
            and men            also
          'and were there also men who'

7  Chie: =un soo  ano ↑a-  o↑toko no ↓hito wa    |ne:,
         mm yeah uhm  oh-  the.men          as.for FP
        'mm yeah uhm  oh- as for the men,'
                                          |((looks up toward
                                             ceiling))
```

visual invocation of the category {hula dancers} to which "the beautiful women" belong

M enquires about men using *mo*, with the implication that men may also be members of the category {hula dancers}.

C excludes men from the category {hula dancers}; with *wa*, C sets in motion **a search procedure** for an appropriate MCD which has {hula dancers} as a category.

```
8 → Mari: [nanka hi:  nanka [tsukechattari >nanka kon'nakoto yatteru wake?<
           like  fire like   lighting       like  like.this  doing   FP
          'were ((they))like lighting the fire, like going sort of like this?'
9 → Chie: [a- ha: >(   ) [a-  hi:  nanka tsukechatteru no.<
           oh  DF          oh- fire like  lighting      FP
          'oh- ha: >(   ) [oh- ((they))'re like lighting the fire.'
```

line 8: M includes men in co-class category {fire-lighters} and simultaneously defines a new MCD: "types of roles in a Hawaiian show".
line 9: C endorses M's categorization operations.

In line 1 Mari encourages elaboration by enquiring about the beautiful women, with a turn-beginning of the form "referent (the beautiful women) + particle *ga*." As Ono et al. (2000) have shown, *ga* is regularly used to foreshadow a forthcoming description of the state of a referent. This prompts Chie to provide a description: "yeah, were illuminated by spotlights and (.) going like this." (lines 2 and 4), portraying their state by enacting the arm motions of hula dance. In other words, through mobilization of grammar and visual conduct, Mari and Chie are characterizing "the beautiful women" as {hula dancers}.

Mari next proceeds to enquire about the men, marking a new reference formulation *otoko no hito mo* "the men" with the particle *mo*, which can potentially be heard as enquiring if the men were also doing hula dance—tantamount to including the men in the just invoked category {hula dancers}. That Chie finds the use of *mo* problematic here is revealed by what happens next (line 7). After embarking on what sounds like an agreement, Chie stops mid-turn and produces a "change-of-state" token, (Heritage, 1984), ↑*a*- "oh-," which may be used to initiate repair (Schegloff, 1992, p. 1305). She then continues by replacing Mari's use of *mo* with *wa*:

```
6→  Mari: [<de  otoko no hito mo:=
           and  men           also
          'and were there also men who'

7    Chie: =un soo  ano ↑a- o↑toko no ↓hito wa    |ne:,
           mm yeah uhm oh- the.men        as.for FP
          'mm yeah uhm oh- as for the men,'
                                                  |((looks up toward
                                                     ceiling))
```

In other words, Mari's uptake in line 8 displays an understanding that Chie's marking of "the men" with *wa* is projecting a turn trajectory that excludes "the men" from the category {hula dancers}. But the *wa*-marked reformulation (line 7) indicates that Chie is simultaneously projecting something more; that Mari goes on to propose that the men may be lighting the fire attests to the fact that *wa* has apparently set in motion a "search" procedure' for a possible category—containing "the men"—which is a co-class category of {hula dancers} in some overarching MCD. Indeed, the anticipatory completion (line 8) evidences that the search has yielded the category {fire-lighters} which is a co-class category of the category {hula dancers} within a larger MCD. Although not made explicit, Mari's mention of fire-lighters evokes an MCD such as "types of roles in a Hawaiian show," which would contain {fire-lighters} as a co-class category of {hula dancers}. Such an understanding on the part of Mari is ratified by Chie through the latter's acceptance and partial repetition of Mari's suggestion in overlap (line 9).

An important factor enabling Mari's anticipatory completion in line 8 is arguably whether the amount of contextual information accumulated up to that point has reached a certain threshold level, thereby providing a reasonable basis for projection. In retrospect, the participants' collaborative work in invoking the category {hula dancers} and including "the beautiful women" in that category (lines 1–4) is analyzable as constituting vital preliminary steps for eventually evoking the larger MCD—"types of roles in a Hawaiian show." Of course, whether an occasion arises for such immanent MCDs to be actively invoked is contingent on how the interaction unfolds. Here, precisely such

How this replacement is treated by Mari can be observed in the overlap that ensues (lines 8 and 9). First, on the basis of Chie's now revised marking of "the men" with *wa*, and without the benefit of hearing how Chie's turn develops, Mari embarks on an anticipatory completion to request confirmation that the men may instead have had a different role such as that of lighting a fire (line 8), thereby registering that she had mistakenly assumed the men were also hula dancers:

```
8 → Mari: [nanka  hi:   nanka  [tsukechattari >nanka kon'nakoto yatteru wake?<
           like  fire  like    lighting       like  like.this  doing   FP
          'were ((they)) like lighting the fire, like going sort of like this?'
9 → Chie: [a-  ha: >((   ))    [a-  hi:  nanka tsukechatteru no.<
           oh  DF              oh- fire like  lighting      FP
          'oh- ha: >((    ))   [oh- ((they))'re like lighting the fire.'
```

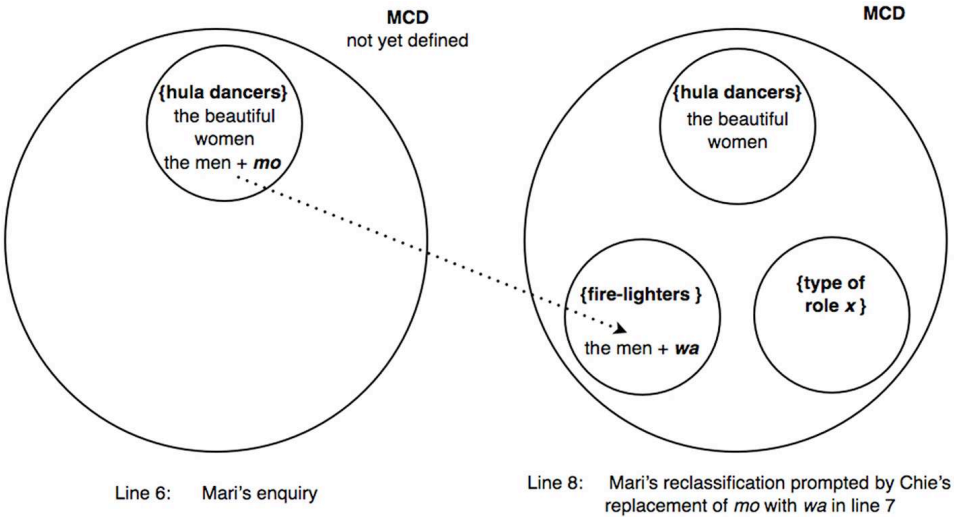


FIGURE 3 | Excerpt (6): Mari marks "the men" with *mo* to enquire if they belong to {hula dancers} (left); subsequently, Chie's marking of "the men" with *wa* prompts Mari to exclude them from {hula dancers} and to reassign them instead to {fire-lighters} (right). MCD on right = "types of roles in a Hawaiian show" = [{hula dancers}, {fire-lighters}, {type of role *X*},...], where *X* is unspecified. Only three of the possible roles have been represented in the figure.

an occasion is presented through Mari's further categorization activity to attempt to classify "the men" (line 6), taken even further by Chie's projected reclassification (line 7), synergistically thrusting the immanent MCD into the scope of interactional relevance. **Figure 3** represents the reclassification resulting from Chie's replacement of *mo* with *wa* in line 7.

A final factor contributing to the anticipatory completion are productional features of Chie's turn-beginning in line 7. Chie displays an attempt to search for a description of {the men} at the end of line 7 partly through her upward glance toward the ceiling suggestive of a word-search, as well as the sound-stretch on the final particle *ne* which can be heard as a move to gain time. Such disruptions in progressivity provide "unprojected opportunities" for Mari to implement an anticipatory completion (see Lerner, 1996), but may have simultaneously given Mari just enough time to execute the cognitive operations made relevant by Chie's production of *wa*.

Incidentally, the fact that Chie targeted the particle *mo* (projecting "sameness") in line 6 for replacement with the particle *wa* (projecting a "contrast") in order to repair Mari's original suggestion that the men may also be engaging in hula dance, bears witness to Chie's understanding of the "inappropriateness" of using *mo* when talking about a referent *otoko no hito* "the men" supposedly not belonging to a previously invoked category {hula dancers}. To rearticulate, *wa* was used not only to exclude "the men" from the category {hula dancers} but also to enable the inclusion of "the men" in another category, which is a co-class category of the overarching MCD "types of roles in a Hawaiian show," within the complement of the category {hula dancers}. This instance contributes toward further buttressing the potentially mutually exclusive nature of the two particles *mo* and *wa* (see Takeuchi, 1999, p. 133), and the capacity of *wa* to mobilize a search procedure for an appropriate MCD.

Consider another instance, this time of a preemptive response, which sheds further light on the operations set in motion by *wa*, the significance of prior contextualizing work, and productional features, as well the ways in which they work in tandem to permit coparticipants to form a basis for projecting the likely trajectory of a *wa*-marked turn-beginning and to respond to it. Furthermore, this instance will be used to demonstrate that the process of anticipating the kind of MCD being invoked may be vastly simplified when the *wa*-marked reference formulation projects an opposite co-class category—i.e., narrowing down the choice to just one candidate co-class category.

In this conversation [same as the one from which excerpt (1) was taken], a group of female students at university were asked to talk freely about their preferences regarding men. The participants have been discussing their likes and dislikes, exemplifying their opinions by referring to members of a popular Japanese, all-male band, including Masa and Shun, who are also topicalized in the excerpt itself. Shortly before the stretch of conversation shown below, the talk had revolved around types of eyebrows and the thicknesses of hair in men, with F expressing a dislike for certain types of eyebrows in men. E then commented that she was disinterested in the types of eyebrows men have, to which L agreed. D nevertheless went on to express her dislike for thick eyebrows, with which F agreed by citing Masa as an example. D took this further by asserting her aversion to men with thick hair. Then, in line 1, L playfully objects to everyone using the band members as exemplars of the traits.

(7) [Sakura 07, 262, Thick or thin features]

```
1    L:   ((smile voice))
          zenbu XXX     ni tatoe[ru  no yame↑te kure↑na:i?
          all  ((band)) P  exemplify N stop     give.NEG
         'can you stop using XXX as exemplars of all ((the traits))?'
2  All:                        [((  clapping  and   laughter  ))

3    L:   [((residual laughter))
4    D:   [nanka saa nanka  Masa   kun  toka saa,
          like   FP  like  ((name)) SFX  e.g. FP
         'like you know, like Masa for instance,
          you know'

5         kao  mo   koi   shi  saa
          face also thick CONJ FP
         'his face is also thick and'

6         nanka ke   toka mo   ko    soo       [jan?
          like  hair e.g. also thick seemingly  FP
         'like his hair also seems thick doesn't it?'
7    F:                                        [a hah hah ha.

8    E    soo na  n  daroo ne:?
          so  COP N  COP   FP
         'that's probably the case,
          isn't it?'

9    D:   mayu[ge   kara shite?.
          eye.brows from do
         'his eyebrows as starters.'
10    F:      [he

11    F:   eh heh heh heh heh

12    D    nanka Shun   kun  wa::  [nanka:
           like ((name)) SFX as.for like
          'as for like Shun        like'
13 → F:                             [usu soo=
                                     thin seemingly
                                   '((his)) seem
                                    thin.'

14   D:   =usu  soo       janai
           thin seemingly  COP
          '((his)) seem thin,
           don't you think?'

15   E:   usui usui zettai.    a↑shi toka nai yo. tabun.   anmari:.
          thin thin absolutely legs   e.g. NEG FP  probably not.much
         '((his)) are thin, absolutely.  as for ((his)) legs
          for example, probably ((he)) hardly has any.'
```

Presents Masa as an example.

Expands description to other "thick" features of Masa, such as his face and his hair.

Describes Masa's thick eyebrows as one example of his thick hair, i.e., Masa ∈ {men with thick features}

Marks new referent Shun with *wa*, thereby putting him in {men with thin features}, a co-class category of the already invoked category {men with thick features}.

Shun ∈{men with thin features} ratified

L's tongue-in-cheek plea to the others (line 1) makes explicit her judgment that the members of the band are being used as exemplars embodying the various attributes of the target population "eligible men." L's objection notwithstanding, D proceeds (line 4) to illustrate her earlier mentioned aversion by citing Masa as embodying thick features—thick (prominent) face (line 5), thick hair (line 6), and thick eyebrows (line 9). By enumerating a range of features of Masa which epitomize the quality of "thickness," D in effect, invokes and makes relevant the category {men with thick features} to which Masa is being assigned. Note that the referent *Masa kun* (line 4) is marked with a topic particle *toka* used here to link Masa to the emerging category as one out of an unspecified number of incumbents (a usage similar to that of *tte* as detailed in Section Basic Categorization/Set-theoretic Actions Performed by Topic Particles).

Having laid the groundwork for further categorization work by making relevant the category {men with thick features} containing Masa, D then names another band member, Shun, through a *wa*-marked reference formulation (line 12). As soon

as this turn-beginning *nanka Shun kun wa*:: "as for like Shun" (line 12) can be heard, F enters with a preemptive response *usu soo* "((his)) seem thin" (line 13), which is built on the prediction that D is projecting exclusion of Shun from the category {men with thick features}. F's response is ratified by D herself through repetition (line 14), and followed by an upgraded agreement and further elaboration by E in line 15. Lines 13–15 exhibit three participants' shared understandings that *Shun kun wa* locally projects a characterization of Shun as having features which are "thin" in some sense, i.e., that Shun belongs to the category {men with thin features}. This instance further substantiates the role of *wa* to assign a referent it marks to a co-class category {men with thin features} of an already invoked category {men with thick features} within an overarching MCD which partitions the universe of discourse (i.e., "eligible men" in this example).

As with the previous excerpt, the preemptive action here is triggered and propelled by a constellation of factors. In addition to the unprojected opportunity for turn-entry created by a sound stretch on *wa*::, F's preemptive response is facilitated through extensive categorization activity prior to line 12, as detailed above. The explicit inclusion of Masa in the category {men with thick features} both prior to the beginning of the extract and in the extract itself establishes a firm foundation for categorizing additional members. What distinguishes this example from excerpts (2) and (6) is that while there was a potentially unspecified number of categories comprising the MCDs invoked in (2) and (6), the category {men with thick features} in the present example utilizes a binary opposition of thick vs. thin, thereby making relevant a MCD comprising two opposing categories and no others. Thus, the situated deployment of *wa* within an interactional environment in which the category {men with thick features} has previously been invoked serves as a ready mechanism for invoking the one and only possible co-class category—{men with thin features}. The resultant MCD is schematized in **Figure 4**.

As discussed previously, two different roles of *wa* have been identified by Maruyama (2003), namely, for performing opposite *or* parallel contrasts. It can be seen here that opposite contrast is the only type of contrast possible when a MCD is defined with reference to some binary opposition. Alternatively, some types of MCDs such as "types of medical doctors" inherently have many co-class categories, in which case, *wa* may trigger a selection from among the potentially multiple parallel categories rather than just one. In this sense, the examples inspected so far suggest that there is a higher order of generality that subsumes both roles under a single operation.

In extracts (6) and (7) above, a *wa*-marked reference formulation triggered an anticipatory completion or preemptive action that was quickly ratified by coparticipant(s). An inspection of the earlier talk revealed that crucial groundwork had already been laid through "adequate" preliminary categorization activities, including implicit or explicit invocation of some category and a candidate MCD. Such preparatory work was argued to underpin the formation of an interactional context ripe for further categorization work. It should come as no surprise, then, that a subsequent reference formulation marked with *wa* can create a fertile moment for triggering coparticipant anticipatory completion or preemptive action. The next section examines instances where a *wa*-triggered uptake is not ratified by the speaker, and explores how such developments may be linked to factors present in the preceding contextualization work.

Wa Used to Mobilize a "Search Procedure" for a Potential MCD when there is Ambiguous or Minimal Contextual Information

In the sequences examined above, there was little apparent contention among coparticipants with regard to the category and MCD being locally invoked, partly owing to the unequivocal contextualization work performed in prior talk. By way of contrast, the first excerpt to be scrutinized here exemplifies how a *wa*-marked reference formulation may make relevant

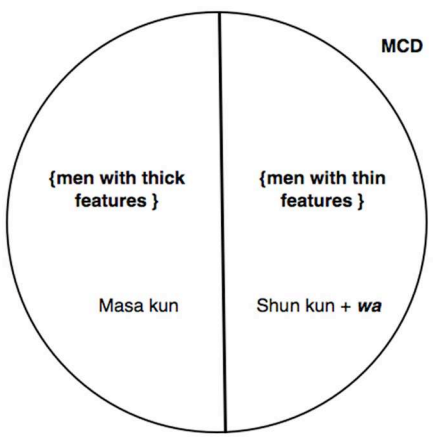


FIGURE 4 | Excerpt (7): The MCD jointly defined by participants by applying a binary opposition partitioning "eligible men" along the parameters "thick" or "thin" features.

multiple possibilities for MCDs due to ambiguities introduced in the immediate interactional environment. Nevertheless, a close tracking of the categorization work undertaken can reveal that participants display concord with respect to the kind of cognitive operation *wa* sets in motion. The final excerpt demonstrates that *wa* may trigger a preemptive action even when it is preceded by little or no preliminary categorization activity, suggesting that participants may resort to general cultural knowledge or "background expectancies" (Garfinkel, 1967) to furnish an independent basis for contextualization.

The following excerpt is from the same conversation as the one from which excerpt (6) was taken, which transpired when Mari and her daughter visited the home of a family friend Chie and her son Ken. Although too lengthy to show here, the categorization work within the excerpt can be understood against the backdrop of points raised earlier in the conversation, as outlined below in sequence:

a. Ken has complained about having had little alternative but to be attentive to others' needs around the house and to be diligent with the housework (i.e., Ken ∈ {attentive people}).
b. Ken has attributed his predicament to the fact that all the women around him (including Chie's close friend Kazuyo who often comes to stay at the house) are purportedly *suekko* "babies of the family," further describing them as *noonoo to suru* "carefree" or "indolent" and completely reliant on Ken to serve them without themselves lifting a finger (i.e., Kazuyo ∈ {carefree people}).
c. Mari has commented that men must nevertheless find such women utterly *kawaii* (i.e., lovable, sweet, cute, endearing, etc.).
d. Chie then portrayed her friend Kazuyo as someone who has little self-awareness that everyone around Kazuyo may find her behavior bothersome (i.e., everyone around Kazuyo ∈ {people who find carefree behavior bothersome}).

(8) [YKH 1 34'23"-34'47"] Kazuyo's husband

```
1    Chie: min↓na   ga  iyagaru  tokoro o    miru to
           everyone  SUB bothered manner OBJ  see  if

2          ↑ya        na  n↓  da  yo,  kitto  t     tsutte
           bothersome COP N   COP FP  surely  QUOT say

3          ya(h)tta  (h)n(h)  da(h)ke(h)do[(hoh)  ↑ne?
           did       N       CONJ                 FP
          '"seeing as though everyone seems bothered,
          they surely must find ((your behavior)) bothersome",
          ((I)) pointed out to her.'
4    Mari:                                [n. ((sniff)) [>de
                                          'mm.           and
5    Chie:                                              [a
                                                        'oh'

6    Mari: Kazuyo  chan wa     goshujin to   ima wa     issho:   desho?=
          ((name)) SFX  as.for husband  with now as.for together COP
          'Kazuyo is now together with her husband, right?'

7    Chie: =u:n.
           mm
          'mm:'

8    Chie:  Kazuyo   chan no  goshujin wa:
            ((name)) SFX  GEN husband  as.for
           'as for Kazuyo's husband'

9 →  Chie:  ka[waii kawaii   Kazuyo  chan na  n °da tte [iu (kara).°
            lovable lovable ((name)) SFX  COP N COP QUOT say because
           '"my lovable, lovable Kazuyo," ((he)) would say,'
10→  Mari:    [mamena    hito                          [ha::↑::::[::↓
               attentive person                          wow
             '((he))'s an attentive/diligent person?'  'wo:::::w'
11   Chie:                                                    [de? (.)
                                                              'and'
```

```
12  Chie: >soo ja  naka ttara Kazuyo  chan okorikuruu kara.<
           so  COP NEG  if   ((name)) SFX  be.furious CONJ
          'otherwise, Kazuyo would be furious'

13        (1.0)

14  Mari: ((lowered pitch)) ha::a↓.=
                            'wo::w'

15  Chie: =un.
          'mm.'
```

In lines 1–3, Chie uses direct reported speech to reenact her attempt to instill in Kazuyo an awareness that everyone must be bothered with her carefree behavior (point *d.* above), using an "extreme case formulation" (Pomerantz, 1986) *min'na* "everyone." Mari then requests confirmation that Kazuyo is "now" with her husband (line 6) (apparently based on prior knowledge of Kazuyo's habit of leaving home, which is explicitly topicalized immediately following the present extract). This is quickly affirmed by Chie (line 7). Mari's move in line 6 can be heard in this specific context as a preliminary to enquiring if it is the husband who does all the housework, which would be contrary to conventional wisdom—an interpretation borne out by the way Mari subsequently performs an anticipatory completion in line 10, as discussed just below. However, before Mari has a chance to articulate the main question, Chie comes in with a new turn-beginning: *Kazuyo chan no goshujin wa:* "as for Kazuyo's husband," by marking "husband" with *wa:* (line 8). As with excerpt (7), the sound stretch on *wa* not only serves as an unprojected opportunity for co-completion but also extends the duration of time for coparticipants to engage in the necessary cognitive operations locally precipitated by the *wa*-marked referent. Indeed, Chie and Mari almost simultaneously go on to complete Chie's turn-beginning. Interestingly, however, their respective turn-continuations are indicative of the invocation of divergent MCDs to partition the social world.

On the one hand, Chie completes her turn with an enactment of how Kazuyo's husband would hypothetically react: *kawaii kawaii Kazuyo chan na n °da tte iu (kara).°* "'My lovable, lovable Kazuyo,' ((he)) would say." (line 9). Given that Chie has just claimed that everyone would be bothered (lines 1–3), to say the husband would find Kazuyo's behavior lovable is to treat the husband as an exception to this rule—i.e., that he would not find her behavior bothersome. In other words, Chie is building on the contextualization work performed by points *c.* and *d.* in choosing a MCD that partitions the population into two categories, by assigning the husband to the category {people who find carefree behavior lovable} in the co-class category of {people who find carefree behavior bothersome}.

On the other hand, Mari's anticipatory completion "((he))'s an attentive/diligent person?" to characterize the husband (line 10) indicates that Mari has appropriated the slot made available by Chie's turn-beginning and pursued the main question projected by her own preliminary query in line 6, and has accordingly partitioned the same population differently. Mari puts the husband in the category {attentive people} which can be seen to be a co-class category of the previously invoked category {carefree people}, thereby orienting to a characterization of the husband which takes into account the prior contextualization work undertaken in points *a.* and *b.* Namely, Ken's earlier complaint about the women around him has made immanent the category {carefree people}, to which he has assigned Kazuyo, as well as the co-class category {attentive people} in which he has already included himself. Mari is now actively invoking these categories (which has until then only been immanent) triggered by Chie's deployment of *wa* (line 8). In sum, whereas Mari is dividing up the universe of discourse into a MCD consisting of opposing categories of attentive vs. carefree people, Chie can be observed to be orienting to the MCD defined by *reactions* to carefree behavior—consisting of opposing categories of {people who find carefree behavior bothersome} and {people who find carefree behavior lovable}. In other words, the concurrent completions by Chie and Mari in lines 9 and 10 respectively index and implement underlying cognitive operations that divide up the population in different ways. The categorization activities performed by Chie and Mari are schematized in **Figure 5**.

The above example illustrates how prior talk can sometimes make relevant multiple MCDs or ways of classifying the larger population. Indeed, if potential ambiguities are introduced by competing dimensions along which to categorize the population in prior contextualizing work, a situated *wa*-marked reference formulation may trigger disparate collaborative completions representing divergent projections of possible turn-trajectories. On a deeper level, however, the excerpt demonstrates that the apparent differences result from the implementation of the same basic cognitive operation mobilized by *wa* on empirically different MCDs. In this sense, excerpt (8) provides even greater warrant for the proposed operations of *wa*.

Alternatively, *wa* is sometimes occasioned to mark a referent in circumstances where there is minimal prior categorization activity to form a basis for identifying an overarching MCD being invoked. Excerpt (9) explores two further workings of *wa*. First, even where there is little preliminary categorization activity, the marking of a referent with *wa* may nonetheless serve as a trigger for coparticipants to make a "reasonable" guess of the categorization activity involved, by resorting to shared cultural knowledge or "background expectancies" (Garfinkel, 1967). Second, by building on such a prediction, participants can go beyond simply anticipating how a current speaker's turn might develop, and preemptively perform some relevant next action [as in excerpts (1) and (7)].

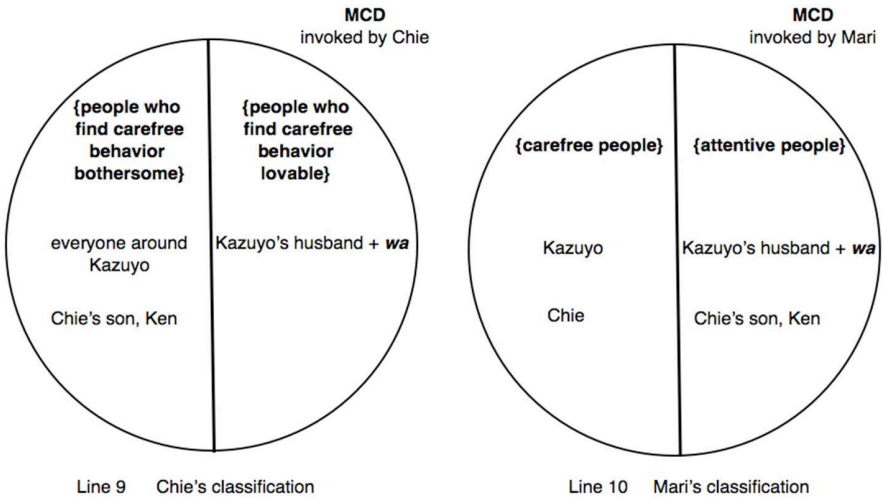


FIGURE 5 | Excerpt (8): Different MCDs triggered by Chie's marking of "Kazuyo's husband" with *wa*. Chie invokes the MCD partitioned by *opposing reactions* to carefree behavior {people who find carefree behavior bothersome} vs. {people who find carefree behavior lovable} **(left)**, whereas Mari invokes the MCD defined by the binary opposition {carefree people} vs. {attentive people} **(right)**.

Japan is often described as a country where there is a persistent normative expectation to get married (to legally tie the knot) by a certain age, even though the average age at first marriage continues to rise (National Institute of Population and Social Security Research, 2011, Table 1-1, p. 2). The following excerpt from a reunion of members of a university yacht club (three women in their late twenties) presents a vivid commentary on the social and personal pressures that may drive one into marriage, even in spite of oneself. Aya, who is the only one out of the three who is already married, has just admitted to the others that her marriage was partly a result of an unremitting buildup of pressure making it difficult to go against the tide.

(9) [Mfriends 2685] Pressure to get married

```
1     Yae: nanka saikin   sa:
           like  recently FP
           'like recently, you know,'

2     Rei: 'n::
           'mm::'

3     Yae: >nanka-< so:nna yoo na (.) fun'iki  ga  dete ki- hajime te  sa:.
           like-    that.kind.of      mood     SUB begin.to.emerge and FP
           'like- that sort of mood of expectation ((e,g., pressure to marry))
            has begun to emerge, and, you know,'

4     Yae: hajimete   sa:.
           first.time FP
           'for the first time, you know.'

5     Rei: ↑a- ho:nto:?
           oh- rea:lly:
           'oh- rea:lly:?'

6     Yae: <dakedo atashi wa    [nanka:
           but     I      as.for like
           'but as for me, like,'
7 →   Rei:                     [↑e-   mukoo      ka[ra:?
                                what  other.side from
                               'what? from the other side?'
                                ((i.e., the partner's side))
```

```
8 →  Aya:                                                  [mukoo      ga?
                                                            other.side SUB
                                                           '((it))'s the other side?'
                                                           ((i.e., the partner's
                                                             side))
9      Yae: 'n::.
            'mm:'

10           (.)

11     Aya: [↑a- demo mawari ga  min'na   soo     [na  no ka ne
             oh- but  around SUB everyone that.way COP FP QP FP
            'oh- but, everyone around one are all being that way, perhaps?'
12     Yae: [nanka demo soo      iu-               [>soo   soo<
             like  but  that.way say                 right right
            'but,  like, that kind-                 >right, right<'

13           mawari ga  soo      dakara:.
             around SUB that.way because
            'because everyone around one is being that way,'

14     Rei: 'n:[n
            'mm:'
15     Yae:    [tabun    soo  iu  kibun de [moriagatteru n da  to   omou [no:.
                probably that say mood  P   worked.up    N COP QUOT think FP
              '((they)) are probably getting all worked up in that kind
               of mood, ((I)) think.'
16     Rei:                                [aaa.                        [n::.
                                           'oh.'                        'mm::'
```

In lines 1 and 3, Yae begins a "second story" by reporting that the pressure for her to marry has likewise gained momentum: "like recently, you know, like- that sort of mood of expectation ((e.g., pressure to marry)) has begun to emerge, and, you know, for the first time, you know." Rei treats this announcement as newsworthy in line 5 by employing a "change-of-state token" ↑*a-* "oh-" and pursues the informing: ↑*a- ho:nto:*? "oh- rea:lly?" (see Heritage, 1984). Yae then resumes her telling in line 6: *dakedo atashi wa nanka*: "but as for me, like," using the contrastive conjunction *dakedo* "but," which adumbrates a contrast, as well as marking *atashi* "I" with *wa*. Notably, this turn-beginning results in an immediate preemptive reaction from Rei: ↑*e- mukoo kara*:? "what? from the other side?" (line 7) containing ↑*e-* "what?" which, as noted previously, is regularly used to mark an informing as departing from one's expectation, supposition, prior knowledge or other orientation (Hayashi, 2009). In other words, without hearing how Yae's turn develops, Rei infers from Yae's marking of "I" with *wa* (line 6) that it is "the other side" (i.e., the partner's side) and not Yae herself who is the source of the pressure. Aya displays a similar understanding through her uptake in line 8: *mukoo ga*? "((it))'s the other side?" (i.e., the partner's side).

Drawing on the discussion so far on the role of *wa*, the marking of *atashi* "I" with *wa* (line 6) would be expected to trigger a search for a category from which "I" would be excluded, by retrospectively searching for some contextualization work in Yae's prior talk. In the excerpts examined previously, the marking of a referent with *wa* was preceded by prior categorization activity that participants could draw upon—such as the invoking of some category and a member of the category. In contrast, there is little if any prior categorization activity in the present excerpt, apart from the mention of the emergence of a mood of expectation that can potentially form the basis of defining a category such as {people creating mood of expectation}.

In the absence of adequate contextual information, the coparticipants appear to base their subsequent categorization activities on background expectancies. The fact that Rei and Aya both identify *mukoo* "the other side" (i.e., the partner's side) as the source of the pressure suggests that the search procedure may have proceeded roughly along the following lines:

a. On reexamining Yae's prior talk, the coparticipants locate the category {people creating mood of expectation}, though Yae has not specified any member of the category.
b. The appearance of *dakedo* "but" and the marking of "I" with *wa* (line 6) can be used to exclude Yae from the category {people creating mood of expectation}, thereby implying that Yae $\in$ {people not creating mood of expectation}.
c. The binary opposition in step *b.* leads to a search for specific person(s) who may be the source of the mood in an overarching MCD.
d. Based on background expectancies and conventional wisdom that there are only two parties to a marriage (i.e., Yae and her partner), the coparticipants select the MCD "parties to a marriage" consisting of two categories {ego's side} and {partner's side} which is "duplicatively organized," i.e., that the set of categories define a social unit (Sacks, 1986, p. 334).

e. The coparticipants appropriate the MCD identified in step *d.* above, and superimpose the structure of this MCD (a binary opposition) over the MCD identified in steps *a.* – *c.* above in order to discover the source of the mood of expectation. As it has already been established (in step *b.* above) that Yae∈{people not creating mood of expectation}, the coparticipants arrive at the conclusion that Yae's partner is the source of the mood—i.e., assigns the partner to the category {people creating mood of expectation} (lines 7 and 8).
f. Consequently, all responsibility for exerting the pressure to marry is attributed to Yae's partner.

It appears that the coparticipants have not only anticipated the trajectory of Yae's turn-beginning in line 6, but have implicitly built on it to initiate their preemptive reactions in lines 7–8.

There is, nevertheless, little guarantee that a "search" will necessarily be endorsed by the original speaker, and "(o)f course, using that procedure for finding the category, you may never come across occasions for seeing that it's 'incorrect'" (Sacks, 1992, Vol. I, p. 337). However, in excerpt (9) an occasion to (in)validate the coparticipants' choice of MCD is afforded. But first, it should be noted that the reactions of Rei and Aya in lines 7–8 contain a potentially problematic inference that the partner may be pressuring Yae to get married against her will. Perhaps in order to counter such an inference, Yae simply proffers a minimal acknowledgement (line 9) followed by a micro-pause (line 10), hearable as implicating some interactional trouble. Indeed, just as Yae begins in line 12 to produce a potential disagreement using the connective *demo* "but" (Mori, 1999), Aya simultaneously comes in (line 11) to treat the minimal response as pointing to a problem with the presumptive inferences drawn earlier by Rei and herself in lines 7 and 8 respectively. In other words, Aya locates the problem as one involving a failed search for an appropriate MCD in the previous turns, i.e., the invocation of the device, "parties to a marriage." This is partly evidenced by Aya's modified formulation in line 11, which is a renewed attempt at searching for another, more "suitable" MCD: she begins with a change-of-state token ↑*a-* "oh-" followed by the activation of an alternative MCD, "everyone around one." Whereas the previous MCD "parties to a marriage" was sharply defined through a binary opposition, the new MCD is diffuse and blurs the earlier distinction between the two parties to marriage—for instance, whether it includes just the couple, their immediate family members, a still wider circle of relatives, friends and acquaintances of the families, or for that matter, even shading into the amorphous notion of *seken* "society at large."

One consequence of invoking this new device is to drain away some of the responsibility for creating the mood of expectation from the partner, and to redistribute it among a broader and fuzzier collection of people. The revised MCD (and the resultant redistribution of responsibility) is now ratified enthusiastically by Yae herself: *>soo soo< mawari ga soo dakara:. tabun soo iu kibun de moriagatteru n da to omou no*: ">right, right< because everyone around one is being that way, ((they)) are probably getting all worked up in that kind of mood, ((I)) think." (lines 12–13 and 15). Interestingly, Yae's talk diffuses the source and nature of the mood of expectation even further, and is rendered highly tentative through expressions such as *omou no* "((I)) think," *tabun* "probably," and the use of anaphoric expressions *soo dakara* "being that way" and *soo iu kibun* "that kind of mood," thereby articulating a sense of ambivalence about the elusive yet pervasive societal pressure to get married. Rei also shows recognition and acceptance of the reformulated MCD in line 16. All told, the newly defined, diffuse MCD "everyone around one" jointly reformulated by Aya and Yae is ultimately endorsed by all three participants. The series of categorization activities performed in this extract is schematized in **Figure 6**.

To summarize, Excerpt (9) exemplifies the deployment of *wa* in an interactional environment preceded by minimal categorization work. Even in such instances, the marking of a referent with *wa* may set in motion a search procedure for a category containing the referent and an overarching MCD. Where there is little prior categorization work to serve as a basis for the search, participants may consult their cultural knowledge and background expectancies as a basis for implementing the cognitive operations of *wa*.

Concluding Comments

This article investigated the potential of *wa* to propose, trigger, and propel anticipatory completions and preemptive action trajectories within locally emergent frames of interaction incorporating interlinked membership categorization activities. The mobilization of *wa* is often preceded by earlier classification activities such as assigning a member of a population to some category. Marking a referent with *tte* represents a typical method to explicitly invoke a new category while simultaneously proposing membership of the referent in that category. More generally, an explicit or implicit invocation of a category and a member of the category through *some* means can create an interactional environment that makes salient extended opportunities for subsequent interlocking categorization activities, which are regularly performed through the differential use of *mo* and/or *wa*. Specifically, while *mo* is used to include another referent in a category that has already been invoked, the data indicate that marking a referent with *wa* indexes a cognitive operation to exclude the referent from an already invoked category and to assign it instead to a contrastive co-class category in a relevant membership categorization device. Detailed examination of instances where the situated marking of a referent with *wa* leads to anticipatory completions and preemptive actions yielded evidence that participants draw on such underlying categorization operations to project the trajectory of the turn-in-progress and to plan a relevant next action.

The projective potential of *wa* has been explored in a range of interactional contexts. First, when progressive groundwork is laid through preliminary contextualization work, participants can develop an increasingly firm basis on which to mobilize the capacity of *wa* to pick out a co-class category from a relevant MCD, and achieve consensus as to how to classify a *wa*-marked

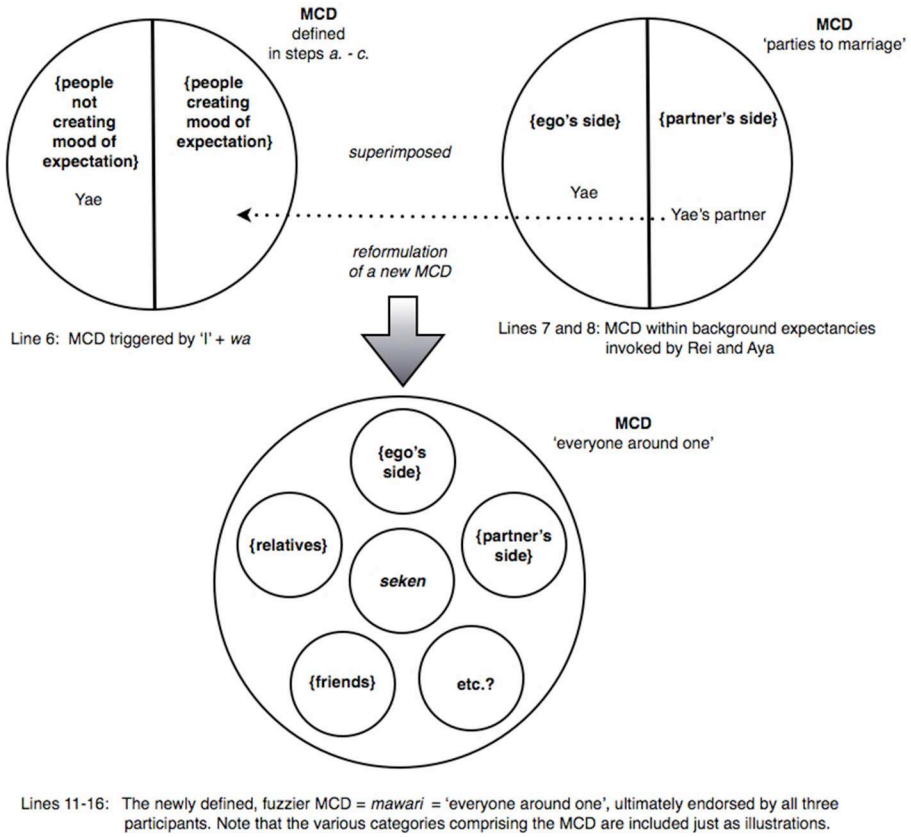


FIGURE 6 | Excerpt (9): Yae's marking of "I" with *wa* triggers a MCD consisting of categories {people not creating mood of expectation} and {people creating mood of expectation}, with Yae belonging to the former (top left). Rei and Aya superimpose a MCD retrieved from their background expectancies in order to discover the source of the mood of expectation to marry **(top right)**. However, Aya and Yae reformulate the just proposed MCD into a fuzzier MCD, thereby diffusing the source of the mood (bottom).

referent. Further, the proffering of a *wa*-marked referent is routinely accompanied by a hitch in progressivity through a sound stretch on *wa* or through the use of fillers such as *nanka*, extending the duration of time available for cognitive processing as well as providing "unprojected opportunities" for entry into the turn-space of the current speaker (see Lerner, 1996). I have argued that the categorization operations implemented by *wa*, together with preparatory contextualization work and temporal-productional features may reach critical mass, and trigger coparticipant anticipatory completions and preemptive actions.

On the other hand, where potential ambiguities are introduced through the immanence of multiple MCDs in the immediate interactional environment, a *wa*-marked referent may engender the relevance of disparate MCDs, representing divergent ways of partitioning members of a population. Nevertheless, inspection of the categorization operations coparticipants perform through *wa* can paradoxically indicate that they are implementing an identical cognitive operation, albeit on different MCDs. Such instances can serve as "deviant case analysis" to further warrant the proposed role of *wa*. Finally, even in cases where no category or MCD has been explicitly invoked in prior talk, the marking of a referent with *wa* may sometimes set in motion a search procedure for a possible MCD containing a category from which the referent is excluded. When there is only minimal contextualization work to draw on within the immediate interactional environment, participants may resort to cultural or background knowledge such as relevant "standardized relational pairs" in order to presumptively identify a likely MCD. The basic categorization operations identified in this study are outlined as algorithms in **Table 1**.

The picture of *wa* which emerges here is as a resource deployed to assemble together a myriad of features in the moment-by-moment unfolding interactional environment toward activating and projecting a specific type of categorization activity, which can compensate for the tendency toward delayed projectability in Japanese conversation (see Tanaka, 1999, 2000). If one were to grant that this portrayal can serve as a realistic model of the actual workings of *wa*, it should be apparent that an enquiry that limits consideration to written or non-interactional data would be unable to capture the extent of the complex processes it points to. The operations enabled by *wa*, which have been a subject of an agelong debate in linguistics, appear to exhibit a remarkable order of systematicity

TABLE 1 | Algorithms.

Algorithm 1	If *y* is a member of a population and *Y* is a category (or a description of a category), then "y *tte Y*" can invoke the category *Y*, and propose the incumbency of *y* in *Y* (denoted $y \in Y$).
Algorithm 2	For a member of a population *y* and a category *Y*, if it has been established that $y \in Y$, then the subsequent marking of another member of the population *x* with *mo* "x *mo*" can assign *x* to the same category *Y* (i.e., $x \in Y$).
Algorithm 3	For a member of a population *y* and a category *Y*, if it has been established that $y \in Y$, then the subsequent marking of another member of the population *x* with *wa* "x *wa*" can exclude *x* from the category *Y* and simultaneously propose the existence of another category *X* to which *x* belongs, *and* a membership categorization device *M* in which *X* and *Y* are co-class categories (i.e., *X* is in the complement of *Y* in *M*).
Corollary to Algorithm 3	As a special case of Algorithm 3 above, if a category *Y* has been defined in such a way as to set up a binary opposition, then the membership categorization device *M* proposed will consist of only two categories *Y* and *X*, where $X = \sim Y$ (i.e., *X* is equal to the complement of *Y*)
Algorithm 4	For a member of a population *x*, if *wa* has been used to mark *x* but no membership categorization device has been implicitly or explicitly specified, then "*x wa*" may activate a "search procedure" to identify a membership categorization device *M* containing categories *X* and *Y* such that $x \in X$ and *Y* is a co-class category of *X* in *M*.

when investigated *in situ* through the lens of conversation analysis and membership categorization/set theory. In this regard, anticipatory completions and preemptive actions offer an indispensable vehicle to catch such processes "in flight," as they provide coparticipants' online commentary on the cognitive processing through which an upcoming trajectory of a turn is being projected and acted upon in the middle of the turn. Particularly revelatory are collaborative completions where two participants concurrently display how they are processing and analyzing one and the same *wa*-marked turn-beginning [such as excerpt (8)]. The fact that the completions occur simultaneously is proof that their respective projections were arrived at independently.

The capacity of "topic particles" has often been cited as a characteristic and prominent feature of the Japanese language to grammatically distinguish a "topic" of discourse from the grammatical subject (see Kuno, 1973; Maynard, 1981, 1987; Hinds et al., 1987; Shibatani, 1990; Noda, 1996; Iwasaki, 2013b), along with some other Asian languages such as Korean and Singaporean English (see Deterding, 2007, p. 61; Leimgruber, 2011). According to Sidnell and Enfield (2012), "some social actions are more readily carried out, or are carried out in specific ways, by speakers of a given language by virtue of the lexicogrammatical properties specific to that languages" (p. 312). As a consequence, the language-specific lexicogrammatical resources used to accomplish particular actions can introduce "collateral effects and in this way give the action a local spin or inflection" (Sidnell and Enfield, 2012, p. 302). The apparently dynamic role of *wa* (and other "topic particles") to project turn-trajectories by implementing categorization activities invites further investigation as a possible "collateral effect" of the lexicogrammatical resources made available in Japanese. Though beyond the purview of this article, a preliminary inspection of the data suggest that participants routinely utilize topic particles for various other, related classifying activities, including negotiating modifications to the definition of a proposed category, adding or deleting members from a category, and mobilizing a search procedure for alternative categories and MCDs, etc. Future cross-linguistic studies on interactional resources used to render visible and analyzable the contingent categorization work oriented to by participants may hopefully serve as stimuli in the exploration of hitherto untrodden terrains of membership categorization activities through comparison of tools available in different languages for engaging in the most human and universal of social actions, namely jointly categorizing the world around us (e.g., Lévi-Strauss, 1969).

Acknowledgments

I would like to express my heartfelt appreciations to Sandy Thompson for all the inspiration and encouragement she has given me and for her enlightening comments on an earlier draft. I am greatly indebted to the reviewers, Makoto Hayashi and Emi Morita, for sharing their deep insights with me in order to strengthen the manuscript. All imperfections that remain are solely my responsibility. My grateful thanks also go to Suzanne Miyata for the use of data from the Sakura corpus (MacWhinney, 2007), as well as to Cambridge University Press and John Benjamins Publishing Company for permissions to adapt figures from Hayashi (2004, p. 350) and Tanaka (1999, p. 155), respectively, for use in **Figure 1**.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01113

References

Cantor, N., and Mischel, W. (1979). "Prototypes in person perception," in *Advances in Experimental Social Psychology Vol. 12*, ed L. Berkowitz (New York, NY: Academic Press), 3–52.

Clancy, P. A., and Downing, P. A. (1987). "The use of *wa* as a cohesion marker in Japanese oral narratives," in *Perspectives in Topicalization: The Case of Japanese Wa*, eds J. Hinds, S. Maynard, and S. Iwasaki (Amsterdam: Benjamins), 3–56.

Couper-Kuhlen, E., and Ono, T. (2007). "Incrementing in conversation: a comparison of practices in English, German and Japanese," in *Special Issue: Turn Continuation in Cross-Linguistic Perspective, Vol. 17,* eds E. Couper-Kuhlen and T. Ono (Pragmatics), 513–552.

Day, D. (2013). "Conversation analysis and membership categories," in *The Encyclopedia of Applied Linguistics*, ed C. A. Chapelle (Oxford: Blackwell Publishing Ltd.), doi: 10.1002/9781405198431.wbeal0211

Deppermann, A. (2011). Notionalization: the transformation of descriptions into categorizations. *Hum. Stud.* 34, 155–181. doi: 10.1007/s10746-011-9186-9

Deterding, D. (2007). *Singapore English.* Edinburgh: Edinburgh University Press.

Egbert, M. (2004). Other-initiated repair and membership categorization—Some conversational events that trigger linguistic and regional membership categorization. *J. Pragmat.* 36, 1467–1498. doi: 10.1016/j.pragma.2003.11.007

Ford, C. E., Fox, B. A., and Thompson, S. A. (2013). "Units and/or action trajectories? The language of grammatical categories and the language of social action," in *Units of Talk—Units of Action*, eds B. Szczepek Reed and G. Raymond (Amsterdam: Benjamins), 13–56.

Fox, B. A., Hayashi, M., and Jasperson, R. (1996). "Resources and repair: a cross-linguistic study of syntax and repair," in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 185–237.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Halmos, P. R. (1960, reprinted 1974). *Naïve Set Theory*. New York, NY: Springer-Verlag.

Hayashi, M. (1997). "An exploration of sentence-final uses of the quotative particle in Japanese spoken discourse," in *Japanese/Korean Linguistics* Vol. 6, eds H.-M. Sohn and J. Haig (Stanford: Center for the Study of Language and Information), 565–581.

Hayashi, M. (1999). Where grammar and interaction meet: a study of co-participant completion in Japanese conversation. *Hum. Stud.* 22, 475–499. doi: 10.1023/A:1005492027060

Hayashi, M. (2000). *Practices in Joint Utterance Construction in Japanese Conversation.* Unpublished Ph.D. thesis, University of Colorado.

Hayashi, M. (2001). "Postposition-initiated utterances in Japanese conversation: an interactional account of a grammatical practice," in *Studies in Interactional Linguistics,* eds M. Selting and E. Couper-Kuhlen (Amsterdam: Benjamins), 317–343.

Hayashi, M. (2003). *Joint Utterance Completion in Japanese Conversation*. Amsterdam: John Benjamins.

Hayashi, M. (2004). Discourse within a sentence: an exploration of postpositions in Japanese as an interactional resource. *Lang. Soc.* 33, 343–376. doi: 10.1017/S0047404504043027

Hayashi, M. (2009). Marking a 'noticing of departure' in talk: *Eh*-prefaced turns in Japanese conversation. *J. Pragmat.* 41, 2100–2129. doi: 10.1016/j.pragma.2008.12.008

Heritage, J. (1984). "A change-of-state token and aspects of its sequential placement," in *Structures of Social Action*, eds J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 299–345.

Hester, S., and Eglin, P. (eds.). (1997). *Culture in Action: Studies in Membership Categorization Analysis*. Washington, DC: University Press of America.

Hinds, J., Maynard, S. K., and Iwasaki, S. (eds.). (1987). *Perspectives on Topicalization: The Case of Wa*. Amsterdam: Benjamins.

Iwasaki, S. (1987). "Identifiability, scope-setting, and the particle *wa*: a study of Japanese spoken expository discourse," in *Perspectives in Topicalization: The Case of Japanese Wa*, eds. J. Hinds, S. Maynard, and S. Iwasaki (Amsterdam: Benjamins), 107–141.

Iwasaki, S. (2008). *Collaborative Construction of Talk in Japanese Conversation*. Unpublished Ph.D. Dissertation, UCLA.

Iwasaki, S. (2009). Initiating interactive turn spaces in Japanese conversation: local projection and collaborative action. *Dis. Proc.* 46, 226–246. doi: 10.1080/01638530902728918

Iwasaki, S. (2011). "The multimodal mechanics of collaborative unit construction in Japanese conversation," in *Embodied Interaction: Language and Body in the Material World*, eds J. Streeck, C. Goodwin, and C. LeBaron (Cambridge: Cambridge University Press), 106–120.

Iwasaki, S. (2013a). "Emerging units and emergent forms of participation within a unit in Japanese interaction: local organization at a finer level of granularity," in *Units of Talk—Units of Action*, eds B. Szczepek Reed and G. Raymond (Amsterdam: Benjamins), 243–275.

Iwasaki, S. (2013b). *Japanese (Revised Edition)*. Amsterdam: Benjamins.

Kaiser, S., Ichikawa, Y., Kobayashi, N., and Yamamoto, H. (2001). *Japanese: A Comprehensive Grammar*. London: Routledge.

Kuno, S. (1973). *The Structure of the Japanese Language*. Cambridge, MA: The MIT Press.

Leimgruber, J. R. E. (2011). Singapore English. *Lang. Linguist. Comp.* 5, 47–62. doi: 10.1111/j.1749-818X.2010.00262.x

Lerner, G. (1991). On the syntax of sentences-in-progress. *Lang. Soc.* 20, 441–458. doi: 10.1017/S0047404500016572

Lerner, G. (1996). "On the 'semi-permeable' character of grammatical units in conversation: conditional entry into the turn space of another speaker," in *Interaction and Grammar,* eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 238–276.

Lerner, G. (2004). On the place of linguistic resources in the organization of talk-in-interaction: grammar as action in prompting a speaker to elaborate. *Res. Lang. Soc. Interact.* 27, 151–184. doi: 10.1207/s15327973rlsi3702_3

Lerner, G., Bolden, G. B., Hepburn, A., and Mandelbaum, J. (2012). Reference recalibration repairs: adjusting the precision of formulations for the task at hand. *Res. Lang. Soc. Interact.* 45, 191–212. doi: 10.1080/08351813.2012.674190

Lerner, G., and Takagi, T. (1999). On the place of linguistic resources in the organization of talk-in-interaction: a co-investigation of English and Japanese grammatical practices. *J. Pragmat.* 31, 49–75. doi: 10.1016/S0378-2166(98)00051-4

Lévi-Strauss, C. (1969). *Totemism.* Harmondsworth: Penguin.

Liddicoat, A. J. (2004). The projectability of turn constructional units and the role of prediction in listening. *Dis. Stud.* 6, 449–469. doi: 10.1177/1461445604046589

MacWhinney, B. (2007). "The TalkBank Project," in *Creating and Digitizing Language Corpora: Synchronic Databases Vol. 1*, eds J. C. Beal, K. P. Corrigan, and H. L. Moisl (Houndmills: Palgrave-Macmillan), 163–180.

Martin, S. E. (1987). *A Reference Grammar of Japanese*. Rutland, VT: Tuttle.

Maruyama, A. (2003). Japanese *wa* in conversational discourse: a contrast marker. *Stud. Lang.* 27, 245–285. doi: 10.1075/sl.27.2.03mar

Matsumoto, Y. (1998). "The complementizer *toyuu* in Japanese," in *Japanese-Korean Linguistics, Vol. 7*, eds. N. Akatsuka, H. Hoji, S. Iwasaki, S.-O. Sohn, and S. Strauss (Stanford: Centre for the Study of Language and Information), 243–255.

Maynard, S. K. (1981). The given/new distinction and the analysis of the Japanese particles *-wa* and *-ga*. *Pap. Linguist.* 14, 109–130. doi: 10.1080/08351818109370527

Maynard, S. K. (1987). "Thematization as a staging device in the Japanese narrative," in *Perspectives on Topicalization: The Case of Japanese Wa*, eds J. Hinds, S. K. Maynard, and S. Iwasaki (Amsterdam: Benjamins), 57–82.

Maynard, S. K. (1993). *Discourse Modality: Subjectivity, Emotion and Voice in the Japanese Language*. Amsterdam: Benjamins.

Mori, J. (1999). *Negotiating Agreement and Disagreement in Japanese: Connective Expressions and Turn Construction*. Amsterdam: Benjamins.

Morita, E. (2005). *Negotiation of Contingent Talk: The Japanese Interactional Particles Ne and Sa*. Amsterdam: Benjamins.

Morita, E. (2012). 'This talk needs to be registered': the metapragmatic meaning of the Japanese interactional particle *yo*. *J. Pragmat.* 44, 1721–1742. doi: 10.1016/j.pragma.2012.07.011

Nakamura, K. (2009). *Interactive Negotiation of Perspectives in Japanese: Predicate-Final Structure as a Resource to Organize Interaction*. Ph.D. Dissertation, University of Wisconsin-Madison.

National Institute of Population and Social Security Research. (2011). *Attitudes Toward Marriage and Family among Japanese Singles (The Fourteenth National*

Fertility Survey in 2010). Department of Population Dynamics Research. Available online at: http://www.ipss.go.jp/index-e.asp (Accessed November 5, 2012).

Noda, H. (1996). *Wa to Ga [Wa and Ga], Shin Nihon Bunpoo Sensho 1 [Selected Works in Contemporary Japanese Grammar 1]*. Tokyo: Kurosio Shuppan.

Ochs, E., Schegloff, E. A., and Thompson, S. A. (eds.). (1996). *Interaction and Grammar*. Cambridge: Cambridge University Press.

Ono, T., and Suzuki, R. (1992). Word order variability in Japanese conversation: motivations and grammaticization. *Text* 12, 429–445. doi: 10.1515/text.1.1992.12.3.429

Ono, T., and Thompson, S. A. (2003). Japanese *(w)atashi/ore/boku* 'I': they're not just pronouns. *Cogn. Linguist.* 14, 321–347. doi: 10.1515/cogl.2003.013

Ono, T., Thompson, S. A., and Suzuki, R. (2000). The pragmatic nature of the so-called subject marker *ga* in Japanese: evidence from conversation. *Dis. Stud.* 2, 55–84. doi: 10.1177/1461445600002001003

Pomerantz, A. (1986). Extreme case formulations: a way of legitimizing claims. *Hum. Stud.* 9, 219–229. doi: 10.1007/BF00148128

Sacks, H. (1972). "An initial investigation of the usability of conversational data for doing sociology," in *Studies in Social Interaction,* ed D. Sudnow (New York, NY: The Free Press), 31–74, 430–431.

Sacks, H. (1986). "On the analyzability of stories by children," in *Directions in Sociolinguistics: The Ethnography of Communication*, eds J. J. Gumperz and D. Hymes (Oxford: Blackwell), 325–345.

Sacks, H. (1992). *Lectures on Conversation (Volumes I and II)*. Oxford: Blackwell.

Schegloff, E. (1992). Repair after next turn: the last structurally provided defense of intersubjectivity in conversation. *Am. J. Sociol.* 97, 1295–1345. doi: 10.1086/229903

Schegloff, E. (2007a). A tutorial on membership categorization. *J. Pragmat.* 39, 462–482. doi: 10.1016/j.pragma.2006.07.007

Schegloff, E. (2007b). Categories in action: person-reference and membership categorization. *Dis. Stud.* 9, 433–461. doi: 10.1177/1461445607079162

Selting, M., and Couper-Kuhlen, E. (eds.). (2001). *Studies in Interactional Linguistics*. Amsterdam: John Benjamins.

Shibatani, M. (1990). *The Languages of Japan*. Cambridge: Cambridge University Press.

Sidnell, J., and Enfield, N. J. (2012). Language diversity and social action: a third locus of linguistic relativity. *Curr. Anthropol.* 53, 302–333. doi: 10.1086/665697

Suzuki, S. (1995). The functions of topic-encoding zero-marked phrases: a study of the interaction among topic-encoding expressions in Japanese. *J. Pragmat.* 23, 607–626. doi: 10.1016/0378-2166(94)00050-O

Szczepek, B. (2000a). *Formal Aspects of Collaborative Productions in English Conversation. InLiSt—Interaction and Linguistic Structures No. 17*.

Szczepek, B. (2000b). *Functional Aspects of Collaborative Productions in English Conversation. InLiSt—Interaction and Linguistic Structures* No. 21.

Takagi, T. (2001). *Sequence Management in Japanese Child-Adult Interactions.* Unpublished Ph.D. Dissertation, University of California, Santa Barbara.

Takeuchi, L. (1999). *The Structure and History of Japanese: From Yamatokotoba to Nihongo*. London: Longman.

Tanaka, H. (1999). *Turn-Taking in Japanese Conversation: A Study in Grammar and Interaction.* Amsterdam: Benjamins. doi: 10.1075/pbns.56

Tanaka, H. (2000). Turn projection in Japanese talk-in-interaction. *Res. Lang. Soc. Interact.* 33, 1–38. doi: 10.1207/S15327973RLSI3301_1

Tanaka, H. (2001a). Adverbials for turn-projection in Japanese conversation: toward a demystification of the 'telepathic' mode of communication. *Lang. Soc.* 30, 559–587. doi: 10.1017/S004740450100402X

Tanaka, H. (2001b). "The implementation of possible cognitive shifts in Japanese conversation: complementisers as pivotal devices," in *Studies in Interactional Linguistics,* eds M. Selting and E. Couper-Kuhlen (Amsterdam: Benjamins), 81–109.

Tanaka, H. (2005). Grammar and the timing of social action: word order and preference organization in Japanese. *Lang. Soc.* 34, 389–430. doi: 10.1017/S0047404505050141

Tanaka, H. (2013). The Japanese response token *Hee* for registering the achievement of epistemic coherence. *J. Pragmat.* 55, 51–67. doi: 10.1016/j.pragma.2013.02.003

Thompson, S. A., and Couper-Kuhlen, E. (2005). The clause as a locus of grammar and interaction. *Dis. Stud.* 7, 481–505. doi: 10.1177/1461445605054403

Tsujimura, N. (1996). *An Introduction to Japanese Linguistics.* Oxford: Blackwell.

Wlodarczyk, A. (2005). "From Japanese to general linguistics: starting with the 'wa' and 'ga' particles," in *Paris Lectures on Japanese Linguistics,* ed A. Wlodarczyk (Tokyo: Kurosio Shuppan), 27–62.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 June 2015
doi: 10.3389/fpsyg.2015.00797

Word-by-word entrainment of speech rhythm during joint story building

*Tommi Himberg *, Lotta Hirvenkari, Anne Mandel and Riitta Hari*

Brain Research Unit, Department of Neuroscience and Biomedical Engineering, School of Science, Aalto University, Espoo, Finland

Movements and behavior synchronize during social interaction at many levels, often unintentionally. During smooth conversation, for example, participants adapt to each others' speech rates. Here we aimed to find out to which extent speakers adapt their turn-taking rhythms during a story-building game. Nine sex-matched dyads of adults (12 males, 6 females) created two 5-min stories by contributing to them alternatingly one word at a time. The participants were located in different rooms, with audio connection during one story and audiovisual during the other. They were free to select the topic of the story. Although the participants received no instructions regarding the timing of the story building, their word rhythms were highly entrained ($\overline{R}$ = 0.70, $p < 0.001$) even though the rhythms as such were unstable ($\overline{R}$ = 0.14 for pooled data). Such high entrainment in the absence of steady word rhythm occurred in every individual story, independently of whether the subjects were connected via audio-only or audiovisual link. The observed entrainment was of similar strength as typical entrainment in finger-tapping tasks where participants are specifically instructed to synchronize their behavior. Thus, speech seems to spontaneously induce strong entrainment between the conversation partners, likely reflecting automatic alignment of their semantic and syntactic processes.

Keywords: turn-taking, entrainment, word rhythm, mutual adaptation, speech, social interaction

OPEN ACCESS

Edited by:
Judith Holler, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Michael Richardson, Univeristy of Cincinnati, USA
Fred Cummins, University College Dublin, Ireland

****Correspondence:***
Tommi Himberg, Department of Neuroscience and Biomedical Engineering, School of Science, Aalto University, Puumiehenkuja 2, FI-02150, Espoo, Finland
tommi.himberg@aalto.fi

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *04 March 2015*
Accepted: *27 May 2015*
Published: *12 June 2015*

Citation:
Himberg T, Hirvenkari L, Mandel A and Hari R (2015) Word-by-word entrainment of speech rhythm during joint story building. Front. Psychol. 6:797. doi: 10.3389/fpsyg.2015.00797

Introduction

During human social interaction, body movements and behavior synchronize at many levels. This interpersonal coordination can be intentional or unintentional, and it can take many shapes. In conversation, participants' utterance length, vocabulary, and information density, as well as body posture and the use of non-verbal gestures often adapt or match (Condon and Ogston, 1967; Kendon, 1970; Giles et al., 1991; Chartrand and Bargh, 1999; Gonzales et al., 2010). Similarly, continuous rhythmic behaviors can entrain, or converge in phase and period (Pikovsky et al., 2001; Clayton et al., 2004). Such an entrainment to a common rhythm can be seen in music and dance, finger tapping, rocking in chairs or gait when walking side by side (Boker et al., 2005; Repp, 2005; Richardson et al., 2007; Nessler and Gilliland, 2009; Himberg and Thompson, 2011). Entrainment has positive affective consequences (Hove and Risen, 2009; Wiltermuth and Heath, 2009), and while foregrounded in music and dance, timing and entrainment also play important roles in verbal and non-verbal communication (Bavelas et al., 1986; Shockley et al., 2003; Cummins, 2009).

During smooth conversation, turn-taking is accurately regulated between the participants, who thereby can avoid overlap of speech and optimize silence between the turns. To time their own contributions correctly, the participants need to be able to predict the end of their partner's turn. Traditionally, turn-taking is said to be governed by a set of linguistic rules (Sacks et al., 1974),

while more contemporary theories have suggested turn-taking to be driven by entrainment of oscillatory processes (Wilson and Wilson, 2005), and to operate at the level of prosody and timing, rather than linguistic units (Cowley, 1998). Turn-taking is often seen as fundamental in human cognition, even as a species-specific, evolutionary adaptation (Sidnell, 2001). The basic mechanisms of turn-taking are thought to be universal, although different languages somewhat vary in the optimal duration of gaps between turns (Stivers et al., 2009).

Interpersonal entrainment is a result of continuous mutual adaptation, as has been demonstrated in simple hand-tapping tasks performed by two persons (Konvalinka et al., 2010) as well as in dance (Himberg and Thompson, 2011). Such mutual adaptation emerges in live dyadic interaction and can be observed already in infants (Malloch and Trevarthen, 2009). For example, when participants read texts together, their verbal outputs are better synchronized when they are in live interaction than when they co-read with recorded speech (Cummins, 2009). Moreover, partners synchronize their finger-tapping better with other humans than with non-responsive computer partners (Himberg, 2014).

Interpersonal coordination in dyads and groups can either occur by matching behaviors, such as gestures, posture, or vocabulary, or as continuous synchronization (Bernieri and Rosenthal, 1991; Dale et al., 2013). Both types of coordination occur in natural conversations, but from an experimental perspective, both have complications. Behavior matching, although commonly observed in many aspects of conversations (e.g. as imitation of the other person's actions, called "chameleon effect" by Chartrand and Bargh, 1999), occurs intermittently, as the interlocutors do not mirror each other's contributions, but rather interact in a complementary fashion (Abney et al., 2014). Also, the time lags of matching are unpredictable, and can be as long as minutes (Louwerse et al., 2012). Continuous synchronization also occurs during natural conversations, for example, the body sways of the interlocutors synchronize. However, these movements are so small that measuring them requires special sensors, and even then the signal is noisy and the data analysis is complicated (Shockley et al., 2003). To overcome these complications, we used a word game where turns change predictably and often enough, and thus we could measure interpersonal coordination from the speech signals.

Our aim was to experiment on interpersonal coordination using a linguistic task, to contrast with the cognitively less challenging finger-tapping tasks that are the traditional approaches to studying intentional synchrony (Repp, 2005). We aimed at a task that would feel natural and be easy to explain to the subjects and would allow us to measure interpersonal synchronization directly from the speech signals, rather than relying on changes in secondary, oscillatory movements, such as swinging a pendulum or rocking in a chair (Richardson et al., 2005, 2007). Unlike Reich et al. (2014) who looked at pitch synchrony between therapists and clients, we were interested in word timing. We thus asked pairs of participants to create stories word by word, each contributing one word at a time. Since turn-taking occurred after every word, we were able to study word timing in a relatively controlled situation. As Finnish is a highly inflected language, each turn consisted of a meaningful word, rather than a preposition, article etc. that do not exist in Finnish (see Supplementary Information 1). Our participants were seated in separate rooms and connected via either an audiovisual link ("video call") or audio-only link ("telephone call"), allowing us to analyze the relative contributions of auditory and visual cues to speech-rhythm entrainment. The terminology and criteria regarding synchronization and entrainment vary largely in the literature (for a review, see Himberg, 2014, pp. 21–35), but in the present study, by word-rhythm entrainment, we refer to phase-locking of the temporal sequences of word onset times of the two participants.

Methods

Participants, Apparatus, Materials

We studied 18 healthy adults (12 males, 6 females; aged 21–43 years, mean ± SD 27.1 ± 0.6 years), all native Finnish speakers, forming 9 sex-matched pairs. After the course of the study had been explained to the subjects, they gave their written informed consent. The study had prior approval by the Ethics Committee of the Hospital District of Helsinki and Uusimaa.

The data were collected during a two-person magneto-encephalography (MEG) experiment, using a MEG2MEG setup (Baess et al., 2012) but only the behavioral results will be reported here. Participants were seated in separate rooms and, depending on the task condition, they had either an audio-only connection (microphones and headphones), or an audiovisual connection where they could also see a video feed of the other participant in natural size on a projection screen positioned 1 m in front of them.

In our custom-made internet-based communication system, the one-way latency is 50 ± 2 ms for audio signal and 130 ± 12 ms for video (Zhdanov et al., in press). In a pilot dyad, the participants reported they did not notice any lags in either audio or video transmission, and they rated the feeling of presence of their partner at 9 on a 10-point scale. Our participants also reported not to have detected the 80-ms asynchrony between the audio and video inputs during normal conversation that was also included in the setup. This feeling of real-life-like presence of the other person is understandable because the audio and video latencies of our system were well under the limits for smooth conversation (100 ms for audio, 500 ms for video; Jansen and Bulterman, 2013), and even under the limits for more delay-sensitive tasks (60 and 140 ms; Kurita et al., 1994). The asynchrony between the audio and video inputs was within the 130-ms integration window within which auditory and visual speech inputs are considered synchronous, when the auditory input precedes the visual one (Dixon and Spitz, 1980; Vroomen and Stekelenburg, 2011). We therefore considered the transmission latencies of our setup to be negligible for our task, where inter-word intervals were over 2 s.

Procedure

Participants were instructed to construct a story, contributing one word at a time in alternating turns. They were free to select the topic of the story, and no instructions were given regarding

the rhythm or the timing of the words. The experimenter indicated which participant was supposed to start. The stories were about 5 min in duration. Each dyad constructed two stories, one in which they had only an audio connection, and another where they also could see each other on screen. The order of conditions was counterbalanced across dyads. Because of time constraints, two of the nine dyads completed the task in only one of the two conditions, leaving 16 stories to be analyzed.

Analysis

We aimed to quantify the rhythm of speech and the interdependence of word timing both for each single individual and between the participants of a dyad. A total of seven instances of coughing, laughing and interruptions due to not hearing the word were removed from the data.

In speech, the stream of stressed syllables generates the word rhythm (Vos et al., 1995; Scott, 1998). In Finnish, word stress occurs on the first syllable of the word (Iivonen, 1998), and therefore we opted to use word onsets as the basis of our word-rhythm analysis. Word onset and offset times were defined in Matlab 7 (MathWorks) from the 48-kHz audio files as the moments where the sound envelope exceeded the level of background noise during silence. After the detection of the onset and offset times, each sound was labeled manually as a word or a non-word and then transcribed. If the actual word was preceded by an interjection (participant saying e.g. "umm…fishing"), the beginning of the interjection was selected as the onset time, as in many cases the interjection and the word were inseparably merged.

Four different time series were extracted from the word-onset and -offset data (**Figure 1A**): inter-turn intervals (ITIs; times between consecutive word onsets for one speaker), inter-word intervals (IWIs; times between successive word onsets in the joint stream), word durations (DURs), and gap durations (GAPs).

The IWI and ITI time series were converted to phase values, and the concentrations of the resulting circular distributions were used as stability and entrainment measures (see Supplementary Information 2). The stability measure represented the "steadiness" of the consecutive IWIs or ITIs, quantifying how similar each time interval was in relation to the previous one: equally long intervals yielded a phase value of zero, while deviations yielded non-zero values ranging from 1 to 359°.

The entrainment measure, on the other hand, reflected the consistency of the interrelationship between the ITIs of the two participants. It was calculated by measuring where, within one participant's ITIs, the other participant's word onsets occurred. If Participant 1 uttered a word at exactly half way the ITI of Participant 2, the phase value was 180°, with deviations from this anti-phase state ranging theoretically from 0 to 359°. In practice, however, the possible range of relative angles was somewhat narrower (we observed it to range from 14 to 326°), because the participants needed to avoid overlaps (zero relative phase would mean that both participants would start their words simultaneously).

For both stability and entrainment, circular distribution measure $\overline{R}$ (Fisher, 1993, p. 32) and mean angle θ were calculated for each trial, as well as for all the data of the experiment. $\overline{R}$ ranges from 0 (no stability or no entrainment) to 1 (perfect stability or perfect entrainment), and it has previously been used in quantifying individual timing stability and especially synchronicity and entrainment in dyadic and group timing (Himberg, 2006, 2014; Rankin et al., 2009; Lucas et al., 2011).

To statistically evaluate whether the word rhythms in trials were stable and/or entrained, we conducted *V*-tests and Kuiper two-sample tests to see if the observed distributions statistically differed from uniform distributions (Fisher, 1993; Jammalamadaka and Sengupta, 2001). For the entrainment measure, we compared the observed distribution with a uniform distribution from 14 to 326°, corresponding to the range of phase angles that was observed in the study.

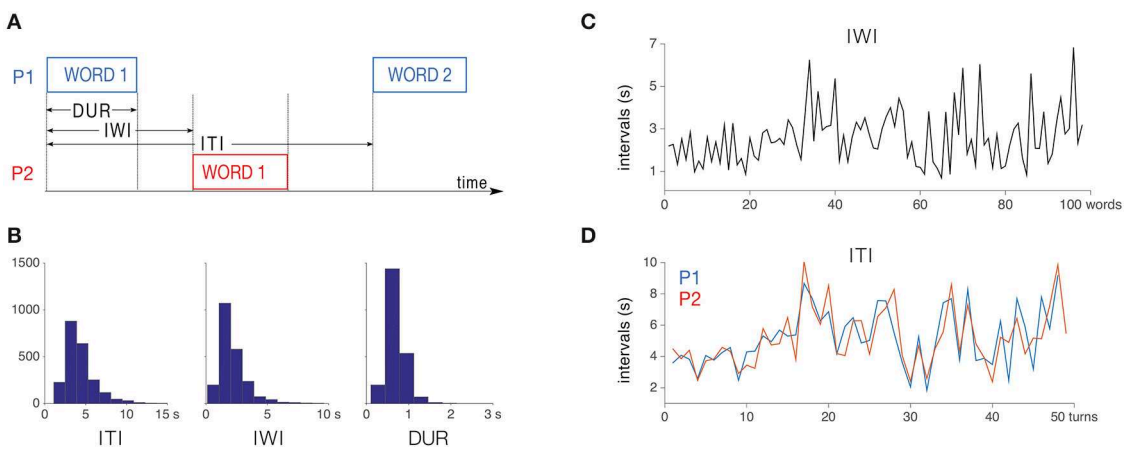


FIGURE 1 | (A) variables extracted from the word onset-offset times; **(B)** histograms of inter-turn intervals, inter-word intervals and word durations; **(C)** IWI, the joint series of word timings, example data from one story; **(D)** ITI, the individual series of word timings, example data from the same story as in **(C)**. Blue line refers to Participant 1 and red line to Participant 2.

Results

General

In the 16 stories by 9 different dyads, a total of 2261 words were uttered, on average 141.3 words per story, or 70.7 (range 43–110) words per participant per story. **Figure 1B** shows the histograms for word durations, IWIs, and ITIs. The word durations were on average (mean ± SD) 0.69 ± 0.23 s, IWIs were 2.14 ± 1.15 s, and ITIs were 4.29 ± 1.80 s.

Participants produced the words in a normal tempo with a mean rate of 3.3 syllables/s, which is comparable to that of normal spoken Finnish (Toivola et al., 2009). As expected, due to the nature of the task, the gaps between words (on average 1.45 s) were longer than in normal, continuous speech (0.5 s, Toivola et al., 2009). The sentences that the participants constructed together were syntactically coherent. Across all pairs, sentences contained on average 9.3 ± 1.7 words (range 2–24), and a story contained on average 14.9 ± 6.8 sentences.

Stability and Entrainment

Figures 1C,D show the ITI and IWI data from an individual story. Both the IWIs and ITIs varied a lot from one word to the next, often by several seconds, making word timing unstable. However, the inter-turn intervals of the two participants (1D) were highly correlated (for this example $r = 0.72$, $p < 0.001$) with each other, indicating high entrainment between the participants.

The circular histograms in **Figure 2** confirm this pattern for the whole experiment, demonstrating that word rhythms were highly entrained even though the individual and joint timings were unstable. The distribution of the relative phase angles (**Figure 2**) calculated from the ITIs has a clear preferred direction toward 180°, indicating anti-phase entrainment. The entrainment measure for the pooled data was $\overline{R} = 0.70$, and $\overline{R} = 0.74 \pm 0.05$ for the 16 individual stories. Instead of varying evenly within its observed range (14–326°), the distribution shows a heavy weighting to anti-phase angles, with 95% of the values concentrated between 78 and 270°. This phase attraction toward the anti-phase was also demonstrated in statistical tests, where, in all stories, the observed distributions deviated statistically significantly from uniform distributions ($p < 0.01$; Kuiper test).

In contrast, the distributions for both the individual word timings (ITI, **Figure 2**) and in the joint time series of word onsets (IWI) were uniformly spread around the circle without any clear preference. The stability measures were very low, $\overline{R} = 0.14$ for ITI, and $\overline{R} = 0.13$ for IWI.

Looking at the 16 joint time series and the 32 individual time-series (16 stories * 2 participants) separately, the stability measure $\overline{R}$ was 0.15 ± 0.064 for the joint timings (IWIs) and 0.18 ± 0.071 for the individual timings (ITIs). *V*-tests confirmed that with the exception of three cases, all individual ITI time-series were unstable, as the distributions did not differ from uniform distribution toward the expected mean direction of zero at $p < 0.05$.

The stability and entrainment scores did not differ between the audio-only and audiovisual conditions ($p = 0.39$ for IWI, $p = 0.15$ for ITI, and $p = 0.15$ for entrainment; paired two-tailed *t*-tests).

Discussion

We found that when two participants were creating stories together, in turns, one word at a time, their word rhythms were strongly entrained. Such a high level of entrainment was unexpected, as the word rhythms themselves were very unstable, and the participants were not given any instructions related to word rhythm, tempo, or timings of their words. Previously, entrainment of comparable strength has been observed in finger-tapping tasks, where the entraining beats occur at equal intervals and the participants are specifically asked to aim for accurate anti-phase timing. The unexpected independence of high levels of entrainment from a stable word-to-word rhythm is in line with the oscillation-based theory of turn-taking (Wilson and Wilson, 2005), which assumes that conversation participants are entrained to a common rhythm that is established by shared syllable timing (Street, 1984). This shared rhythm governs the

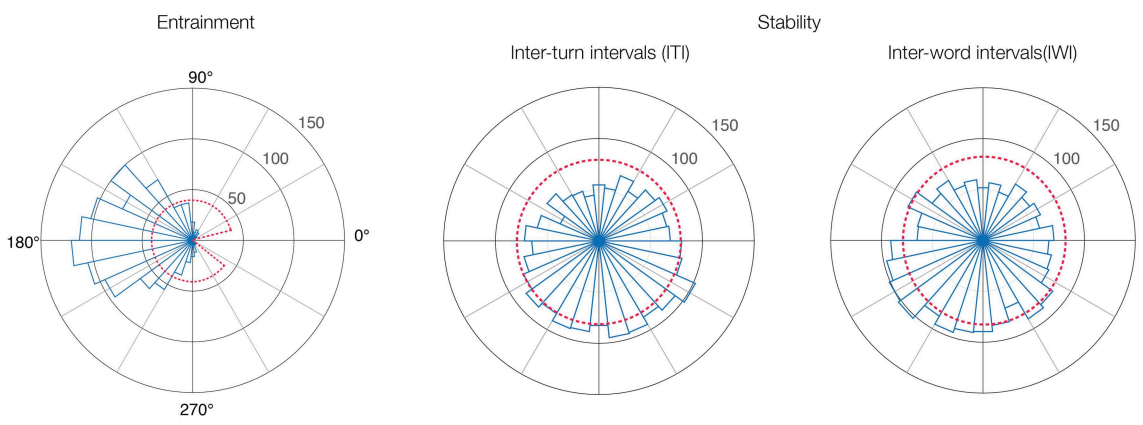


FIGURE 2 | High entrainment in the absence of stability. Circular histograms of relative phase (entrainment) and stability distributions in the whole experiment. Red dashed lines represent uniform distributions of data and the range of observed data.

participants' "readiness" to take turns, and it helps them to optimize turn-taking so that it does not comprise overlaps and long silences.

Interpersonal coordination and adaptation occur in a wide range of tasks, such as pronouncing letters of the alphabet (Kawasaki et al., 2013) or in anti-phase finger tapping (Nowicki et al., 2013). These rather simple tasks mainly recruit automatic entrainment processes, whereas our task of joint story building required advanced cognitive operations to guarantee that the story evolved in a meaningful and smooth manner.

As an automatic and subconscious process, entrainment is assumed to subserve communicative interaction (Gallese, 2001, 2005; Himberg, 2014), and in our task, participants needed to be aligned at the semantic and syntactic levels, as well as the speech-process level (Clark, 1996; Garrod and Pickering, 2004). The high entrainment that we observed could be what allowed the participants to reach this multi-level, multimodal coordination (Dale et al., 2013).

In our study, stability and entrainment were statistically similar in "telephone-like" trials (with only auditory connection between the participants) and "video-call-like" trials (with auditory and visual connection between the participants). This result partly agrees with the results of a previous corpus study of face-to-face as well as telephone dialogs, where pause durations between participants were highly correlated in both types of conversations, suggesting entrainment to a common rhythm even in telephone-mediated conversations (Ten Bosch et al., 2004). However, in the corpus study, the pauses were longer and more variable in the face-to-face conversations. We did not observe such differences, possibly because due to the simultaneous MEG recording, our participants were asked to sit still, which limited the amount and utility of gestural communication between them. All task-critical information was delivered through the auditory channel.

Although our participants conducted the story-building task in a laboratory setting that restricted their body movements, highly entrained speech rhythms emerged spontaneously in their interaction. This mutual adaptation of speech rhythms implies speech as a strong inducer of entrainment, even when the participants just hear each other.

Acknowledgments

This study was supported by the European Research Council (Advanced Grant #232946 to RH), Academy of Finland grants #131483 and #263800, Finnish Graduate School of Neuroscience, and the Doctoral Programme Brain and Mind. We thank Pamela Baess, Mathieu Bourguignon, Mia Illman, Veikko Jousmäki, Jyrki Mäkelä, Jussi Nurminen, Lauri Parkkonen, Petteri Räisänen, Ronny Schreiber, and Andrey Zhdanov for contributing to the realization of the 2-person MEG experiments, during which these behavioral data were collected.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00797/abstract

References

Abney, D. H., Paxton, A., Dale, R., and Kello, C. T. (2014). Complexity matching in dyadic conversation. *J. Exp. Psychol. Gen.* 143, 2304–2318. doi: 10.1037/xge0000021

Baess, P., Zhdanov, A., Mandel, A., Parkkonen, L., Hirvenkari, L., Mäkelä, J. P., et al. (2012). MEG dual scanning: a procedure to study real-time auditory interaction between two persons. *Front. Hum. Neurosci.* 6, 83. doi: 10.3389/fnhum.2012.00083

Bavelas, J., Black, A., Lemery, C., and Mullett, J. (1986). I show how you feel: motor mimicry as a communicative act. *J. Pers. Soc. Psychol.* 50, 322–329. doi: 10.1037/0022-3514.50.2.322

Bernieri, F. J., and Rosenthal, R. (1991). "Interpersonal coordination: behavior matching and interactional synchrony," in *Fundamentals of Nonverbal Behavior,* eds R. S. Feldman and B. Rimé (Cambridge: Cambridge University Press), 401–432.

Boker, S., Covey, E., Tiberio, S. and Deboeck, P. (2005). "Synchronization in dancing is not winner-takes-all: Ambiguity persists in spatiotemporal symmetry between dancers," in *Proceedings of the North American Association for Computational, Social, and Organizational Science* (Notre Dame, IN).

Chartrand, T., and Bargh, J. (1999). The chameleon effect: the perception-behavior link and social interactions. *J. Pers. Soc. Psychol.* 76, 893–910. doi: 10.1037/0022-3514.76.6.893

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clayton, M., Sager, R., and Will, U. (2004). In time with the music: The concept of entrainment and its significance for ethnomusicology. *ESEM Counterpoint* 1, 1–45. Available online at: http://oro.open.ac.uk/id/eprint/2661

Condon, W., and Ogston, W. (1967). A segmentation of behavior. *J. Psychiatr. Res.* 5, 221–235. doi: 10.1016/0022-3956(67)90004-0

Cowley, S. J. (1998). Of timing, turn-taking, and conversations. *J. Psycholinguist. Res.* 27, 541–571. doi: 10.1023/A:1024948912805

Cummins, F. (2009). Rhythm as entrainment: the case of synchronous speech. *J. Phon.* 37, 16–28. doi: 10.1016/j.wocn.2008.08.003

Dale, R., Fusaroli, R., Duran, N., and Richardson, D. C. (2013). The self-organization of human interaction. *Psychol. Learn. Motiv.* 59, 43–95. doi: 10.1016/b978-0-12-407187-2.00002-2

Dixon, N. F., and Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception* 9, 719–721. doi: 10.1068/p090719

Fisher, N. (1993). *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.

Gallese, V. (2001). The shared manifold hypothesis. From mirror neurons to empathy. *J. Conscious. Stud.* 8, 33–50.

Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. *Phenomenol. Cogn. Sci.* 4, 23–48. doi: 10.1007/s11097-005-4737-z

Garrod, S., and Pickering, M. (2004). Why is conversation so easy? *Trends Cogn. Sci.* 8, 8–11. doi: 10.1016/j.tics.2003.10.016

Giles, H., Coupland, N., and Coupland, J. (1991). "Accommodation theory: communication, context, and consequence," in *Contexts of Accommodation: Developments in Applied Sociolinguistics,* eds H. Giles, N. Coupland, and J. Coupland (Cambridge: Cambridge University Press), 1–68.

Gonzales, A. L., Hancock, J. T., and Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communic. Res.* 37, 3–19. doi: 10.1177/0093650209351468

Himberg, T. (2006). "Co-operative tapping and collective time-keeping - differences of timing accuracy in duet performance with human or computer partner," in *Proceedings of the ICMPC 9,* eds M. Baroni, A. R. Addessi, R. Caterina, and M. Costa (Bologna), 377.

Himberg, T. (2014). *Interaction in Musical Time*. Doctoral dissertation, Faculty of Music, University of Cambridge, Cambridge.

Himberg, T., and Thompson, M. R. (2011). Learning and synchronising dance movements in South African songs – cross-cultural motion-capture study. *Dance Res.* 29, 305–328. doi: 10.3366/drs.2011.0022

Hove, M. J., and Risen, J. L. (2009). It's all in the timing: interpersonal synchrony increases affiliation. *Soc. Cogn.* 27, 949–960. doi: 10.1521/soco.2009.27.6.949

Iivonen, A. (1998). "Intonation in Finnish," in *Intonation Systems: A Survey of Twenty Languages,* eds D. Hirst and A. Di Cristo (Cambridge: Cambridge University Press), 311–327.

Jammalamadaka, S. R., and Sengupta, A. (2001). *Topics in Circular Statistics,* Vol. 5. Singapore: World Scientific Publishing.

Jansen, J., and Bulterman, D. C. (2013). "User-centric video delay measurements," in *Proceedings of the 23rd ACM Workshop on Network and Operating Systems Support for Digital Audio and Video* (Oslo: ACM), 37–42.

Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., and Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Sci. Rep.* 3:1692. doi: 10.1038/srep01692

Kendon, A. (1970). Movement coordination in social interaction: some examples described. *Acta Psychol.* 32, 101–125. doi: 10.1016/0001-6918(70)90094-6

Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol.* 63, 2220–2230. doi: 10.1080/17470218.2010.497843

Kurita, T., Lai, S., and Kitawaki, N. (1994). Effects of transmission delay in audiovisual communication. *Electron. Commun. Jpn.* 77, 63–74. doi: 10.1002/ecja.4410770306

Louwerse, M. M., Dale, R., Bard, E. G., and Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cogn. Sci.* 36, 1404–1426. doi: 10.1111/j.1551-6709.2012.01269.x

Lucas, G., Clayton, M., and, Leante, L. (2011). Inter-group entrainment in Afro-Brazilian congado ritual. *Empir. Musicol. Rev.* 6, 75–102. Available online at: http://hdl.handle.net/1811/51203

Malloch, S., and Trevarthen, C. (2009). "Musicality: communicating the vitality and interests of life," in *Communicative Musicality - Exploring the Basis of Human Companionship,* eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 1–11.

Nessler, J. A., and Gilliland, S. J. (2009). Interpersonal synchronization during side by side treadmill walking is influenced by leg length differential and altered sensory feedback. *Hum. Mov. Sci.* 28, 772–785. doi: 10.1016/j.humov.2009.04.007

Nowicki, L., Prinz, W., Grosjean, M., Repp, B. H., and Keller, P. E. (2013). Mutual adaptive timing in interpersonal action coordination. *Psychomusicol. Music Mind Brain* 23, 6–20. doi: 10.1037/a0032039

Pikovsky, A., Rosenblum, M., and Kurths, J. (2001). *Synchronization - a Universal Concept in Non-linear Sciences (No. 12)*. Cambridge: Cambridge University Press.

Rankin, S. K., Large, E. W., and Fink, P. W. (2009). Fractal tempo fluctuation and pulse prediction. *Music Percept.* 26 , 401–413. doi: 10.1525/mp.2009.26.5.401

Reich, C. M., Berman, J. S., Dale, R., and Levitt, H. M. (2014). Vocal synchrony in psychotherapy. *J. Soc. Clin. Psychol.* 33, 481–494. doi: 10.1521/jscp.2014.33.5.481

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992. doi: 10.3758/BF03206433

Richardson, M. J., Marsh, K. L., and Schmidt, R. C. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 62–79. doi: 10.1037/0096-1523.31.1.62

Richardson, M., Marsh, K., Isenhower, R., Goodman, J., and Schmidt, R. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Hum. Mov. Sci.* 26, 867–891. doi: 10.1016/j.humov.2007.07.002

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 696–735. doi: 10.1353/lan.1974.0010

Scott, S. K. (1998). The point of p-centres. *Psychol. Res.* 61, 4–11. doi: 10.1007/PL00008162

Shockley, K., Santana, M.-V., and Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 326–332. doi: 10.1037/0096-1523.29.2.326

Sidnell, J. (2001). Conversational turn-taking in a Caribbean English Creole. *J. Pragmatics* 33, 1263–1290.

Stivers, T., Enfield, N., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Hum. Commun. Res.* 11, 139–169. doi: 10.1111/j.1468-2958.1984.tb00043.x

Ten Bosch, L., Oostdijk, N., and De Ruiter, J. P. (2004). "Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues," in *Proceedings of the 7th International Conference Text, Speech and Dialogue*, eds P. Sojka, I. Kopecek, and K. Pala Brno (Berlin: Springer).

Toivola, M., Lennes, M., and Aho, E. (2009). "Speech rate and pauses in non-native Finnish," in *Proceedings of the 10th Annual Conference of the International Speech Communication Association.* (Brighton).

Vos, P. G., Mates, J., and van Kruysbergen, N. W. (1995). The perceptual centre of a stimulus as the cue for synchronization to a metronome: evidence from asynchronies. *Q. J. Exp. Psychol.* 48, 1024–1040. doi: 10.1080/14640749508401427

Vroomen, J., and Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: not that special. *Cognition* 118, 75–83. doi: 10.1016/j.cognition.2010.10.002

Wilson, M., and Wilson, T. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* 12, 957–968. doi: 10.3758/BF03206432

Wiltermuth, S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5. doi: 10.1111/j.1467-9280.2008.02253.x

Zhdanov, A., Nurminen, J., Baess, P., Hirvenkari, L., Jousmäki, V., Mäkelä, J. P., et al. (in press). An internet-based real-time audiovisual link for dual meg recordings. *PLoS ONE.*

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 May 2015
doi: 10.3389/fpsyg.2015.00509

The effects of processing and sequence organization on the timing of turn taking: a corpus study

*Seán G. Roberts *, Francisco Torreira and Stephen C. Levinson*

Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

The timing of turn taking in conversation is extremely rapid given the cognitive demands on speakers to comprehend, plan and execute turns in real time. Findings from psycholinguistics predict that the timing of turn taking is influenced by demands on processing, such as word frequency or syntactic complexity. An alternative view comes from the field of conversation analysis, which predicts that the rules of turn-taking and sequence organization may dictate the variation in gap durations (e.g., the functional role of each turn in communication). In this paper, we estimate the role of these two different kinds of factors in determining the speed of turn-taking in conversation. We use the Switchboard corpus of English telephone conversation, already richly annotated for syntactic structure speech act sequences, and segmental alignment. To this we add further information including Floor Transfer Offset (the amount of time between the end of one turn and the beginning of the next), word frequency, concreteness, and surprisal values. We then apply a novel statistical framework ("random forests") to show that these two dimensions are interwoven together with indexical properties of the speakers as explanatory factors determining the speed of response. We conclude that an explanation of the of the timing of turn taking will require insights from both processing and sequence organization.

Keywords: turn-taking, processing, sequence organization, frequency, concreteness, surprisal, random forests

OPEN ACCESS

Edited by:
Manuel Carreiras, Basque Center on Cognition, Brain, and Language, Spain

Reviewed by:
Blair C. Armstrong, Basque Center on Cognition, Brain, and Language, Spain
Matthew D. Wood, US Army Engineer Research & Development Center, USA

****Correspondence:***
Seán G. Roberts, Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen, 6525 XD, Netherlands
sean.roberts@mpi.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *30 January 2015*
Accepted: *10 April 2015*
Published: *13 May 2015*

Citation:
Roberts SG, Torreira F and Levinson SC (2015) The effects of processing and sequence organization on the timing of turn taking: a corpus study. Front. Psychol. 6:509. doi: 10.3389/fpsyg.2015.00509

1. Introduction

Imagine a species that squawks at conspecifics. If it only has one message type (signaling e.g., "Here I am"), messages will have low information value. If there is only one rule of use, namely "one at a time," communication will exhibit turn-taking, but not much other sequential patterning. Marmoset communication perhaps come close to this (Takahashi et al., 2013). Human communication differs radically on both dimensions: there is immense complexity on the informational parameter as well as the sequential one (Levinson, 2013b). In this paper we explore how these two parameters conspire to explain the temporal properties of human communication.

The core ecological niche for language use is in conversation: that is where language is learnt and the bulk of it is used. A key property of conversation is that participants take turns at talking. This is a demanding environment for language comprehension and production: So short is the average transition between turns that participants in a conversation must often simultaneously comprehend the current turn and plan the next turn (Levinson, 2013a). This suggests that demands on processing such as low frequency words or turns with dense information (Piantadosi et al., 2011) or more abstract concepts (Walker and Hulme, 1999) should influence the timing of turn transitions.

That is, the duration of gaps between turns may reflect the amount of processing required to comprehend the previous turn and plan the upcoming turn.

Equally, however, conversational language use is characterized by two striking constraints. The first is a turn-taking system which minimizes gaps and discourages overlaps (Sacks et al., 1974); this is at least partially normative (interrupting is after all rude). The second is the mapping of structure across turns: a greeting is responded to with a greeting, a question (preferably) by an answer, an offer by an acceptance or declination, and so forth (Schegloff, 2007). This suggests that the major constraints come from interaction in context, and that the timing of turn-taking is above all sensitive to the constraints of sequence organization (Sacks et al., 1974; Schegloff, 2007). Studies from the field of conversation analysis demonstrate that the timing of turn taking may be sensitive to these constraints. Long gaps (i.e., of more than 700 ms) between turns are generally avoided in part because participants may be competing to take a turn at talk and it is the first speaker who takes the floor that generally keeps it. But in addition delayed turn transitions are interactionally marked in some interactional sequences, especially those in which an initial turn sets up an expectation for a specific type of response, as in questions and answers, offers and their uptake, requests and their compliance, etc. (see Stivers et al., 2009; Kendrick and Torreira, 2015). Hence a long pause after a request can be read as presaging non-compliance (Levinson, 1983). All of this suggests that interactional constraints could be of equal or greater importance for the timing of turn taking than simple processing constraints. Likewise, by the rules of turn-taking, certain types of utterances such as backchannels and repairs do not appear to be subject to the usual turn-taking constraints (i.e., avoidance of overlaps and long gaps) and may appear in overlap or be overlapped more frequently than other types of utterances (see Levinson et al., 2015). In sum, then, turn timing is sensitive to the normative structure of turn-taking and the sequential structure of conversation. Participants do not seem to begin a turn as soon as they have sufficiently processed the prior turn and planned their own turn, but rather hold off speaking until the other has finished their turn. For example, speakers generally identify possible points where a turn transition would be relevant in the interlocutor's turn before launching articulation of their own turn (see Levinson et al., 2015; Torreira et al., 2015; Bögels and Torreira, in press). On the other hand, speakers may begin a turn at talk without having fully planned their turn, by using filled pauses (e.g., "uh," "um") at the beginning of their turn in order to "buffer" their comprehension or planning (Clark and Fox Tree, 2002).

At the same time, it is unlikely that there is no relationship between the duration of turn transitions and cognitive processing requirements. It may simply not be possible to plan and launch an interactionally relevant turn following an extremely long, syntactically torturous sentence spoken extremely quickly. Teasing these two domains apart is not easy. Regarding the processing constraints, effects may be small and measures of such information may be difficult to compute. Real conversations, unlike controlled psycholinguistic experiments, are also subject to a large amount of noise. The ideal dataset would include a wide range of utterance types, but natural conversation is inherently subject to skewed distributions. This means that measures such as the frequency of words in a turn and the length of a turn will often be correlated. In order to get a reasonable sample, a large database of automatically processable conversation is needed. Such a quantitative approach goes rather against the tradition of work in conversation analysis, which is largely qualitative in nature, focusing on specific phenomena observed in close detail. However, in recent conversation analytic work, quantitative measures have increasingly been applied to qualitative coding (e.g., Clayman et al., 2007; Stivers et al., 2009). For example, interesting insights on the time course of language planning during turn-taking can be provided by controlling the sequential interactional context and other contextual relevant variables (e.g., several corpus studies on the timing of turn transitions in question-answer sequences, Stivers et al., 2009; Stivers and Enfield, 2010; Strömbergsson et al., 2013; Torreira et al., 2015). This demonstrates that, while qualitative analysis is often a powerful tool for explaining conversational phenomena, it is also possible to uncover and interpret systematic trends in a quantitative dataset provided that the researcher exerts some degree of control over the relevant contextual factors.

The Switchboard corpus (Godfrey et al., 1992; Calhoun et al., 2010) strikes a reasonable balance between the requirements of the two approaches, from theories of processing difficulty on the one hand, and the careful coding of conversational instances on the other. Tens of thousands of turns have been automatically collected and segmented, as well as hand-transcribed for a range of dialogue acts (e.g., different types of questions, statements, backchannels) relevant to sequence organization (see below). The aim of this paper is to assess to what extent measures of sequential organization on the one hand and cognitive processing on the other can explain the timing of turn taking. We use the statistical framework of Random Forests, explained below, to compare the importance of different variables in the distribution of transition times between turns.

This paper asks the following basic question: does sequence organization matter for the timing of turn taking beyond a battery of processing variables known to affect language processing? More precisely, do measures of sequence organization, albeit the coarse measures that are possible to extract from large corpora, contribute to the explanation of the timing of turn taking over and above measures of cognitive processing?

The amount of data and the number of variables makes the number of individual queries that can be asked of this kind of data very large. Also, as this paper shows, many variables are correlated, making it difficult to assess the strength of a relationship in isolation. By answering the question above and getting a "big picture" impression of the data, we hope to provide a map to fruitful future research.

The next section introduces the phenomenon of turn taking in interaction. Next, some predictions are made regarding how various cognitive processing and sequence organization measures should be related to the timing of turn taking. A short introduction to random forests is given before presenting the methods and results.

2. Turn Taking in Interaction

Conversations take place between two or more speakers who typically take turns at talk, usually minimizing overlapping talk ("overlaps") and long turn transitions without talk ("gaps"). The "floor transfer offset" (FTO) provides a way of measuring gaps and overlaps in one single continuous variable (De Ruiter et al., 2006; Stivers et al., 2009; Heldner and Edlund, 2010). FTO is measured as the duration between the end of one turn and the beginning of another turn for pairs of turns involved in a floor transfer. FTO is negative if the turns overlap and positive if there is a gap between them. Cross-culturally, FTOs appear to be strikingly similar, with mean values ranging from 7 to 468 ms in a diverse sample of 10 languages (Stivers et al., 2009) (this range is small considering that the latency in the planning of a single word is of the order of 600 ms, see Levelt et al., 1999).

This paper focuses on conversations between two speakers. Throughout the paper, we will refer to "T1" as the turn prior to a floor transfer and "T2" as the turn following the floor transfer. We will refer to speaker A as the speaker of T1 and speaker B as the speaker of T2. Note that, in many cases, T2 becomes T1 for the next floor transfer in the conversation. Because of this, not all floor transfers involve the same kind of interactional contingency (e.g., a question and its answer vs. an answer to a question and an unrelated statement opening a new conversational sequence).

There are some previous studies of the distribution of FTOs. For example, Strömbergsson et al. (2013) find that FTOs for question-answer sequences are affected by the type of question asked, the type of response given and the topic of conversation. For example, responses were slower to open questions than wh-questions or polar questions. However, this study did not consider processing factors, analyzed the effects of T1 and T2 independently and was based on linear relationships within a restricted sequence type. Our study uses an order of magnitude more data, a wider range of sequence types and considers properties of both T1 and T2 together.

3. Cognitive Planning and Comprehension

Here we list some measures relevant to either production, comprehension, or both, whose importance we can readily check in the data to hand. We consider a number of hypotheses about how these might play a role in response times, measured in FTO.

3.1. Turn Length

By definition, longer turns can have longer periods of overlap with another turn. Moreover, longer utterances are likely to be more complex than shorter utterances, requiring more processing. However, a longer utterance also gives more time for a listener to begin planning her own turn. Therefore, the predictions for effect of the length of T1 on FTO values are not clear without taking other measures of the content of the turn into account. On the other hand, the prediction for T2 length may be clearer. Planning a long utterance should generally take longer than planning a short one, so the FTO is expected to increase as the length of T2 increases.

3.2. Frequency

Psycholinguistic research has shown that word frequency plays a crucial role in ease of processing, both in comprehension and production. In lexical decision experiments for example (i.e., where participants must decide whether a displayed word is a real word or not, in as short a time as possible), frequent words are responded to more rapidly than infrequent words (Balota et al., 2007). This predicts that turns consisting of higher frequency words should be comprehended and produced faster, therefore reducing the turn transitions in which they are involved.

3.3. Concreteness

Words that refer to concrete entities (e.g., "ball") contrast with words that refer to abstract entities (e.g., "justice"). Concreteness ratings have been shown to correlate with lexical decision times, with concrete words being comprehended faster (Schwanenflugel et al., 1988). Concrete words are also more easily recalled and produced than abstract words (Hanley et al., 2013). This predicts that both T1 or T2 turns with many abstract words may lead to longer gaps between them.

3.4. Surprisal

Surprisal is a measure of the amount of information a word carries about the upcoming words in a phrase. For example, the word "the" gives the listener little information about what the next word might be beyond syntactic category, while the word "helter" is almost certain to precede the word "skelter." Various theories of processing suggest that speakers adapt their utterances to spread out the information in a sentence evenly in order to robustly transmit the signal (Piantadosi et al., 2011). In this context, the inverse of surprisal is also a measure of the "projectability" of turns (Magyari and De Ruiter, 2012) (although not necessarily of turn endings). Surprisal is conceptually the same as cloze probability (i.e., the probability of experimental participants using a word as a completion to a sentence fragment), which is used in many experiments looking at word processing (e.g., Kutas and Hillyard, 1984).

3.5. Syntactic Complexity

Syntactically complex utterances require more processing than simpler ones. Syntactically complex sentences make greater demands on working memory (Kemper and Rash, 1988) and are harder to produce and understand (Kemper et al., 1989).

When responding to a turn, speakers must comprehend the previous turn and plan their own turn. If speakers take longer to comprehend turns with complex syntactic structures than turns with simple ones, then comprehension resources may be diverted from planning the response, making the FTO longer. At the same time, if a speaker wants to produce a complex syntactic structure, this could take more time to plan, also making the FTO longer. The prediction is that FTOs become longer as the syntactic complexity of either T1 or T2 increases.

4. Sequence Organization

Various measures of sequence organization are discussed below.

4.1. Adjacency Pairs

Some types of turn make a response relevant. For example, if T1 includes a question, T2 is expected to provide an answer. Answers, on the other hand, do not make the same kind of demands on the next speaker. Therefore, it is possible to identify turns that have initiating actions, like questions, and turns that have responding actions, like answers. When an initiating action, calling for a specific type of response in next turn, is followed by a relevant responding action, the turns form an adjacency pair.

The predictions about the timing of these types of turns, and whether they appear in a particular combination, are not clear. On the one hand, if initiating actions can be recognized easily, then responding actions may be produced closer to the turn end. This may be possible through the internal design of the turn (Drew, 2013; Levinson, 2013a), or through pre-ambles prior to T1 such as pre-offers (e.g., "Are you doing anything tonight?"), which set the context for initiating an offer such as an invitation. In this case, one would expect the timing of the question following a pre-sequence to be more tightly timed. Also, just as lexical frequency aids processing, so frequent adjacency pairs may be quicker to comprehend or produce. On the other hand, responding actions must "fit" with the previous turn, which may require more planning and therefore delay the response. There may also be no particular requirement in terms of timing for turns that do not form adjacency pairs.

One aspect of adjacency pairs that has been studied in terms of timing is preference (Atkinson and Heritage, 1984). Dispreferred responses, such as declinations to offers, invitations, and requests, are often delayed (Kendrick and Torreira, 2015). Delayed transitions may project the valence of the response and so allow the speaker of T1 to begin planning the third turn (the next T1) immediately (Levinson, 1983, 2013a; Clayman, 2002). For example, a delayed or hesitant response after an offer may be followed by an upgraded offer. For these reasons, although dispreferred responses themselves may be belayed, turns following dispreferred responses may have shorter FTOs.

4.2. Response Tokens

Speakers can signal that they understand what is being said with back-channels or response tokens (Gardner, 2001). These include acknowledgement tokens ("yeah," "mm"), continuers ("mm-hm") and news markers ("oh," "really?," Heritage, 1984). While these are often produced "in the clear" they may appear in overlap without competing for the turn. Continuers, for example, are often overlapped by the prior speaker (Local, 1996; Levinson et al., 2015).

4.3. Laughter

Laughter has a variety of interactional uses beyond signaling joy or humor (Jefferson, 1984; Haakana, 2002; Glenn, 2003). The literature on laughter in interaction demonstrates that although laughter may occupy a turn-like slot (e.g., after a joke), laughing (or a sequence of laughter syllables) is often not treated as competing for the floor in the same way as an ordinary utterance might be, but may be superimposed on it by the speaker or be delivered in overlap by listeners (Glenn, 1989; Ford and Thompson, 1996). The lack of turn organization is indicated by the timing of laughter, which can be targeted at the content of the turn (a "recognition point") rather than turn boundaries (Jefferson, 1974; Glenn, 1989). Therefore, laughter may often occur in overlap. Furthermore, overlapping talk is common in sequences containing laughter when humor is involved, and is not treated as problematic by the speakers. Jefferson (1974) identifies two types of laughter: a speaker may laugh after being "invited" to laugh, for instance by the previous speaker laughing, or a speaker may "volunteer" laughter unprompted. While types of laughter are difficult to code for automatically, the turns that include laughter can be identified in the Switchboard corpus. There are four possible combinations: both T1 and T2 include laughter (T1 "invites" laughter and overlap is possible); only T2 includes laughs ("volunteered" laughter, likely to be at a "recognition point" and therefore can occur in overlap); only T1 includes laughter (T1 "invites" laughter, but T2 does not respond, it is likely that T2 is an ordinary turn after a gap); neither turn includes laughter (an ordinary turn transition, therefore a gap).

5. Interactions between Processing and Sequence Organization

Processing and sequence organization accounts make different predictions for some variables. For example, a faster speech rate in T1 would be predicted to lead to a longer gap due to higher processing demands in the comprehender. In contrast, some theories of turn-timing in Conversation Analysis see timing as rhythmic (Couper-Kuhlen, 1993), and would predict that faster speech rates would lead to shorter gaps.

We note that the constraints of processing and sequence organization may not be entirely disparate mechanisms. For example, Stivers et al. (2009) note that negative answers are slower. This may be because the responder is treating the answer as dispreferred (not in line with the expectation indicated by the polarity of the question), and is therefore proferring it reluctantly. But equally, it is well-known that negative responses are harder to process both in comprehension and production (Clark, 1976). In addition, frequency effects and expectability (or its converse surprisal) may apply to both processing and sequencing constraints. Certain types of turn project other types of turn. Thus, a question in T1 makes it interactionally relevant for T2 to provide an answer. Turn transitions may be shorter between these "adjacency pairs," since adjacency pairs are more predictable and therefore aid comprehension and allow planning to begin sooner. That is, frequent, predictable structures and may aid fast transitions in the same way as frequent words do.

Speakers may overlap with an incoming turn when they wish to signal that they recognize in advance what is about to be said (so called "recognitional overlap," Jefferson, 1986), and in tokens of agreement (Stolt, 2008). While this is an observation from the sequence organization literature, it may be measured by surprisal: words which have a large amount of information about the upcoming words allow prediction of the end of the turn.

If the timing of turn taking is the primary "ecology" to which language has to adapt (Levinson, 2006), certain processing effects may only apply after taking sequence organization factors into account. For example, planning of T2 can often

begin when the pragmatic action of T1 can be recognised (Levinson, 2013a). Action ascription is often independent of syntactic structure, instead being dependent largely on sequential context (Gisladottir et al., 2012). An additional overlap between processing and conversational organization is that the latter makes systematic provision for processing problems. Thus, English makes provision for signaling a small processing hitch (uh) vs. a larger one (um) (Clark and Fox Tree, 2002, see application to the Switchboard corpus in a post by Liberman (2014), http://languagelog.ldc.upenn.edu/nll/?p=14991). Consequently, there may be an asymmetry in the predictions for the syntactic complexity of A's turn and B's turn. While T2 has no way of influencing the relationship between syntactic complexity and when T1 ends (apart from other-initiated repair), there is the option of "buffering" planning at the beginning of T2. Speakers often use turn-preserving placeholders, or hesitation markers, such as "uh" and "um" at the start of their turns to minimize the gap between turns. They may use this extra time to plan their response. This asymmetry in the options for T2 predicts that the syntactic complexity of T2 would only be correlated with the FTO when excluding initial parts of T2 that were simply turn-preserving placeholders.

In summary, the timing of turn taking may be heavily context dependent. In this case, we would not expect linear effects of processing measures over the whole data, nor simple categorical effects of sequence organization across the board. Instead, we would expect some relationships to be evident only in certain conditions. Typical regression approaches to statistical modeling are not effective at exploring this kind of data. Because of this we use a random forests framework, which can discover context-dependent relationships.

6. Materials and Methods

Conversations were taken from the Switchboard corpus (Godfrey et al., 1992), a large corpus of telephone conversations recorded in the United States of America in the 1990s. Participants who did not know each other were connected by an automatic switchboard and were assigned a topic of conversation, which was automatically recorded. The corpus has been annotated on different levels over the years since its first release. In this study we use several layers of annotations as compiled in the NXT-Switchboard Corpus (Calhoun et al., 2010).These include segmentation of phonetic segments and words in time, which can be used to estimate the duration of turns at talk and the floor transfer between turns. Due to a flaw in the original data collection, the timing of part of the corpus is unreliable (see Calhoun et al., 2010). For this reason, recordings with unreliable timings were discarded in our study. Utterances have been hand-annotated for dialogue acts, such as yes/no questions or backchannels (Jurafsky et al., 1997). Words are annotated for parts of speech and organized into syntactic trees (Marcus et al., 1999). There is also meta-data on the speakers such as age, sex and location in the USA. Obviously, visual cues are not present in this dataset.

We processed the Switchboard files using specifically designed software (Lubbers and Torreira, 2014). This extracted the FTO between turns (Section 6.1). We categorized the dialog acts of each turn into sequence organization categories and identified turns with laughter and dispreferred responses (Sectio 6.2). For each turn in the database, we also calculated various measures of processing, such as frequency, surprisal, and concreteness, and used the syntactic annotations from the Switchboard corpus to estimate syntactic complexity (Section 6.3).

6.1. Calculating Floor Transfer Offset

The corpus provides timing segmentation of phonological words (originally segmented by Deshmukh et al., 1998). We approximated "turns" by "gluing" phonological words together if they were from the same speaker and had less than 180 ms gap between them. The floor transfer offset (FTO) or "gap" and "overlap" duration between turns from different speakers was calculated using the same method as in Heldner and Edlund (2010). Transitions involving very long gaps or overlaps were discarded from the analyses (FTOs lower than -2200 ms or above 2200 ms, less than 2% of the final data). The distribution of FTOs fits well with distributions reported in other studies (see Section 7).

FTOs were also re-calculated, ignoring T2 initial turn-preserving placeholders, so that we can report FTOs with and without initial hesitation markers. These were identified as in Strömbergsson et al. (2013), as the tokens "uh," "um," and "well." An alternative coding was done with identification based on the syntactic category of the initial word being an interjection, filler or discourse marker (the category "UH" from Calhoun et al.'s coding). We calculated the FTOs from the end of T1 to the beginning of the first word in T2 which was not a turn-preserving placeholder. For this set of data, T2s that consisted of only turn-preserving placeholders were excluded.

6.2. Sequence Organization Data

The Switchboard corpus is annotated with dialog acts (Jurafsky et al., 1997). These are similar to speech acts, but include categories suited for spoken conversations such as backchannels. These dialog acts were grouped into sequence types: first pair parts, second pair parts, opening and closing sequences, backchannels, repairs or "other" (see **Table 1**). For each dialog act type, a set of dialog acts was identified which would make a well-formed adjacency pair. For example, a yes/no question projects a yes or no answer.

Laughter is coded in the Switchboard transcripts, sometimes as a separate feature, and sometimes within the orthographic transcript. Turns that included laughter were identified. Preferred and dispreferred responses were identified with similar criteria as in Kendrick and Torreira (2015). Transitions where T1 initiates a question were identified (with tags "open-q," "tag-q," "wh-q," "yn-q," "yn-decl-q," "commit"). Within these, any T2 that included an accepting dialog act ("affirm," "yes," "answer") were marked as preferred responses, while all others were marked as dispreferred responses. The frequency of every possible pair of dialog acts surrounding an FTO was extracted. Obviously, the measures above are coarse approximations of the qualitative judgments of conversation analysts. However, they are useful for getting a general picture of how the principles of sequence organization could interface with principles of processing.

TABLE 1 | The NXT dialog act categories and how they map onto sequence organization types.

NXT category	Description	Expected next categories	Initiating	Responding	Response token	Valence
decl-q	Declarative Wh-Question	answer,statement	Y			
open	Conventional-opening		Y			
open-q	Open-Question	neg,affirm,no,yes,statement,reject	Y			
or	Or-Clause	neg,affirm,no,yes,statement,reject	Y			
repeat-q	Signal-non-understanding		Y			
sum	Summarize/Reformulate		Y			
tag-q	Tag-Question	neg,affirm,no,yes,statement,reject	Y			
wh-q	Wh-Question	answer,statement,reject	Y			
yn-q	Yes-No-Question	yes,no,affirm,neg,statement	Y			
yn-decl-q	Declarative Yes-No-Question	yes,affirm,statement	Y			
acknowledge	Response Acknowledgment			Y	Y	
backchannel	Backchannel			Y	Y	
backchannel-q	Backchannel as question			Y	Y	
ans-dispref	Dispreferred answers			Y		Neg
hedge	Hedge			Y		Neg
maybe	Maybe/Accept-part			Y		Neg
neg	Negative non-no answers			Y		Neg
no	No answers			Y		Neg
reject	Reject			Y		Neg
affirm	Affirmative non-yes answers			Y		Pos
agree	Agree/Accept			Y		Pos
answer	Other answers			Y		Pos
yes	Yes answers	accept		Y		Pos
apprec	Appreciation			Y		
abandon	Abandoned or Turn-Exit					
apology	Apology	agree,downplay				
close	Conventional-closing	close				
commit	Offers, Options, and Commits					
completion	Collaborative Completion					
directive	Action-directive					
downplay	Downplayer					
excluded	Excluded - bad segmentation					
hold	Hold before response					
opinion	Statement-opinion	agree,opinion,disagree,accept				
other	Other					
third-pty	3rd-party-talk					
quote	Quotation					
repeat	Repeat-phrase	agree				
rhet-q	Rhetorical-Questions	agree				
self-talk	Self-Talk					
statement	Statement-non-opinion	statement				
thank	Thanking	downplay				
uninterp	Uninterpretable					

6.3. Linking the Switchboard to Processing Measures

The turns were linked to various measures of processing. Utterance length was measured in syllables, as included in the NXT-Switchboard corpus. We calculated speech rate using the method from Wightman et al. (1992). This calculates the departure from the expected duration, calculated from the sum of mean phone durations in the corpus.

We estimated word frequency from the Switchboard corpus itself. The count of each word for each part of speech in the transcript of the full corpus was taken (the same method as Potts, 2011, except we also automatically removed tense and number inflection from nouns and verbs in order to improve

the frequency estimates). The full Switchboard corpus includes around 15 million tokens. For each turn, the mean frequency of words was calculated. Larger corpora give estimations of frequency that better predict processing measures such as lexical decision times (e.g., the Subtlex corpus estimates, Brysbaert and New, 2009), but estimates are also sensitive to genre, for which the Switchboard is by definition a good match. In any case, the source of frequency estimates did not affect the general results (see Supplementary Materials 1).

Words from each turn were lemmatized and linked with concreteness ratings from a large ratings study (Brysbaert et al., 2014), matched for part of speech. A measure of surprisal was taken from Piantadosi et al. (2011), which is based on the amount of information a word contains about the following words in the Google n-gram corpus of English. For each turn, we extracted the surprisal value for each word and calculated the mean surprisal value for the turn. In addition, we estimated the uniformity of the information density by taking the mean deviation from the expected uniform information density over words.

We estimated syntactic tree depth from the NXT-Switchboard syntactic trees. The depth of a tree is the maximum number of nodes between the root and any tip in the tree. The maximum depth of any tree in a turn was taken as the maximum depth for that turn. We also measured the number of clauses in each turn, calculated as the number of "S" sentence nodes in all trees of the turn.

Altogether 19,754 turn transitions were found for which each of the 30 predictor measures were available. These came from 348 conversations involving 231 speakers, totaling around 31 h of conversation. The vast majority of the conversations lasted between four and a half and five minutes, as specified in the instructions given to participants. Speakers produced an average of 12 FTOs per minute.

6.4. Random Forests

This paper aims to contrast measures of processing with measures of sequence organization in the explanation of turn transitions. However, many of the considered variables are highly correlated. This can invalidate the assumptions of a typical regression approach (the estimates of individual effects are unstable and the standard errors inflate, leading to misleading comparisons between the strengths of individual predictors and an under-estimation of significance of individual effects). As reported below and in the Supplementary Materials, many of the independent variables in the Switchboard data are correlated.

One solution to this problem is to use the method of "random forests" (Breiman, 2001). This is an approach based on regression (and classification), though the analyses are not linear regressions across the whole data. Instead, a "binary decision tree" (also called classification and regression tree or recursive partitioning, Strobl et al., 2009) uses the predictor variables to split the data into sub-sets. However, the structure of a decision tree is not robust to the selection of variables or sub-sets of data. In order to overcome this problem, many trees are run with sub-sets of predictor variables (hence a random "forest"), then the findings are aggregated to determine the relative importance of different variables.

First, the concept of a decision tree is reviewed. A decision tree is a hierarchy of yes/no-questions that splits data into sub-sets. To illustrate this, consider the tree in **Figure 2**. This was generated with FTO as the dependent variable and four measures of sequence organization (whether T1 includes an initiating action, whether T2 includes a responding action, whether T1 includes laughter and whether T2 includes laughter). For clarity, only the first three levels are shown.

The data is divided at each node of the tree, and the leaves of the tree show the mean FTO for that sub-set of the data in a bar chart. Above each bar chart is a number labeled *n* which represents the number of observations in that sub-set. The tree can be read like a solution to a game of "20 questions." If you are asked to guess the value of an FTO, the decision tree aims to show you the optimal sequence of yes-no questions that will guide your guess. The tree can also be read like a set of rules that describe patterns in the data (e.g., in **Figure 1**, "if the turns form an adjacency pair, the FTO will be a short gap, unless there is invited laughter, in which case the FTO will be in overlap.")

The first decision is whether T1 includes an initiating action (e.g., a question). For a given turn transition, if T1 is initiating, then we follow the right branch. The next "question" splits the data into T2s with responding actions (e.g., answers) and those without. If T2 does include a responding action, we follow the branch to the left, and are asked whether T1 included laughter. If not, then we end up at a terminal category which we might label "adjacency pair," summarized in a bar chart. This bar chart indicates that the mean FTO is around 200 ms, based on 1130 samples (agreeing well with other studies, e.g., Stivers et al., 2009).

Every turn transition can be assigned to one of the terminal categories. For example, turn transitions where T1 is an initiating action, but T2 is not a responding action (a kind of dispreferred response) have a mean FTO of around 300 ms. This fits with work showing that dispreferred responses tend to be delayed (Kendrick and Torreira, 2015). On the other side of the tree, the questions split the data up into whether there is laughter in T1 or T2. Invited laughter, when there is laughter in T1 and T2 produces a mean FTO of around −150 ms (overlap). Again, this is in line with the literature on laughter (see above).

The algorithm that generates the tree works as follows. First, the strength of association between each predictor variable and FTO is determined by a statistical test of independence. The variable with the strongest association is chosen as the first node in the tree. The data is divided according to this variable into two sub-sets. The process repeats recursively with each sub-set until all predictor variables are statistically independent from FTO in each leaf of the tree.

The tree in **Figure 1** was generated directly from data using this automatic algorithm, but exhibits many of the empirical observations in the existing literature. Variables used in decisions nearer the top of the tree have a greater influence on the outcome, so the tree would also predict that sequence type is more important than laughter.

However, our data include continuous variables as well as categorical variables. **Figure 2** shows a second tree generated with both sequence organization and processing predictor variables. The first decision is the sex of the speaker of T1. For a given

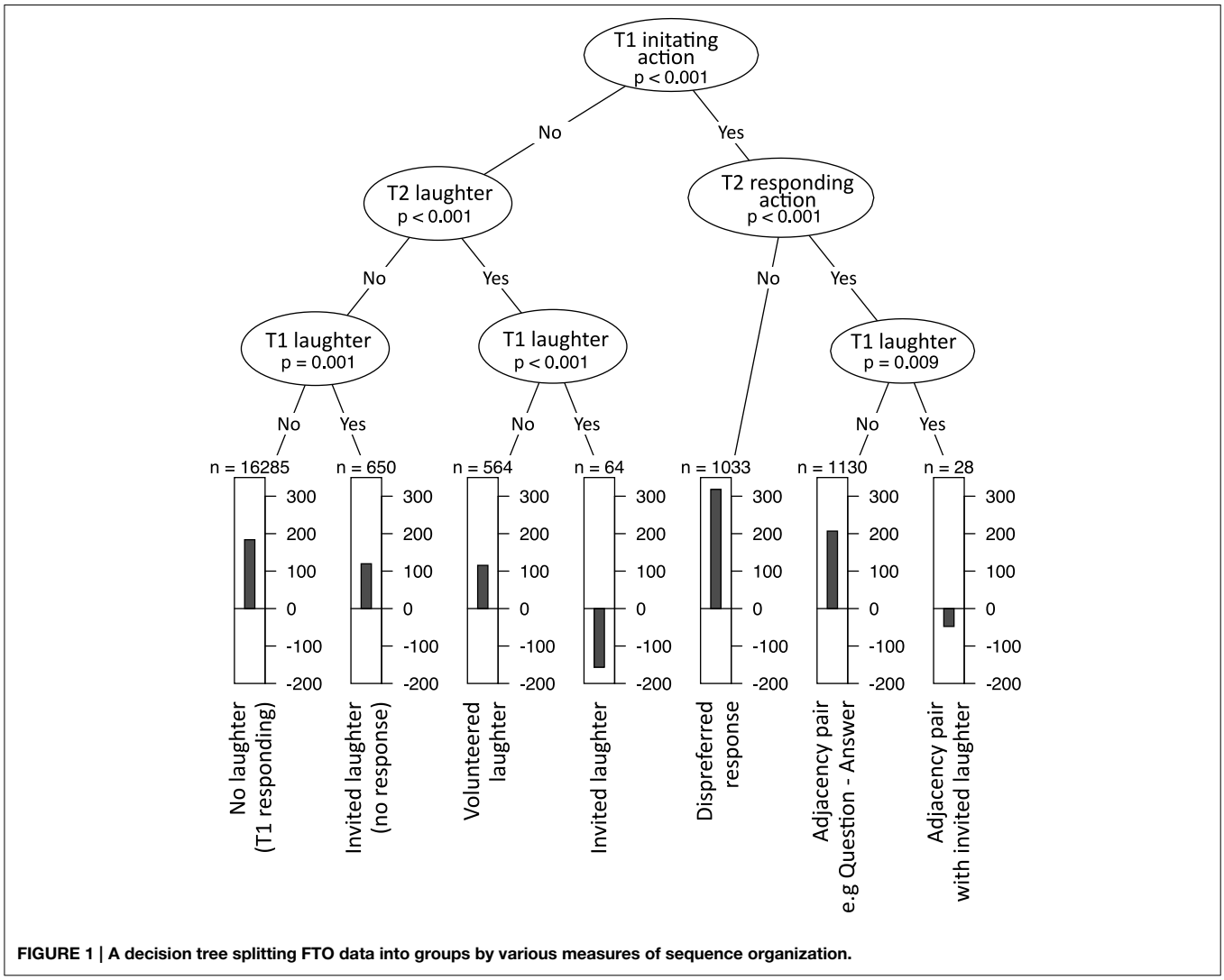


FIGURE 1 | A decision tree splitting FTO data into groups by various measures of sequence organization.

turn transition, if T1 is spoken by a male, then we follow the left branch. The next "question" splits the data into T1s with initiating actions (e.g., questions) and T1s with responding actions (e.g., answers). This continues all the way down the tree, so that the leftmost bar chart shows the mean for FTOs where T1 was spoken by a male, T1 ended with an initiating action and T2 was spoken by a male. Looking at the next bar chat to the right, we see that females have lower FTOs than males when T1 includes an initiating action. For the sub-set with responding actions, we see that the duration of T1 matters, with long turns leading to shorter FTOs than short turns. This goes against the trend in the overall data for long turns to elicit longer gaps. In this way, the decision tree has separated a sub-set of data that behaves differently to the rest, and which helps explain some of the variation.

On the other side of the tree, the second decision is the sex of the speaker of T2. Comparing the leaves on the right, we see that two female talkers tend to produce lower FTOs. Speech rate of T1 is included twice on the next level—the tree cuts the continuous variable at different points for male and female T2 (variables can only be divided into two categories at any one branch, but may be further sub-divided at a later stage). This reflects the trend for males to speak faster than females. For both male and female speakers of T2, slower speech in T1 (higher T1 delta) leads to shorter FTOs. The rightmost leaf represents 314 cases of FTOs between two female speakers where T1 is speaking very slowly (high delta). In this case, the mean FTO is in overlap.

The tree in **Figure 2** shows the first three levels of a full tree. A larger tree based on the full data is available in the Supplementary Materials.

One problem with decision trees is that their structures are not robust. The structure is sensitive to the selection of predictor variables and the particular sample of data (Strobl et al., 2009; Tagliamonte and Baayen, 2012). For example, the choice of the first variable may have been based on a marginal trend in the data, but may have a large effect on the subsequent choices. One way around this problem is to generate a "forest" consisting of a number of randomly generated trees. A sub-sample of the data and a selection of variables are chosen randomly for each

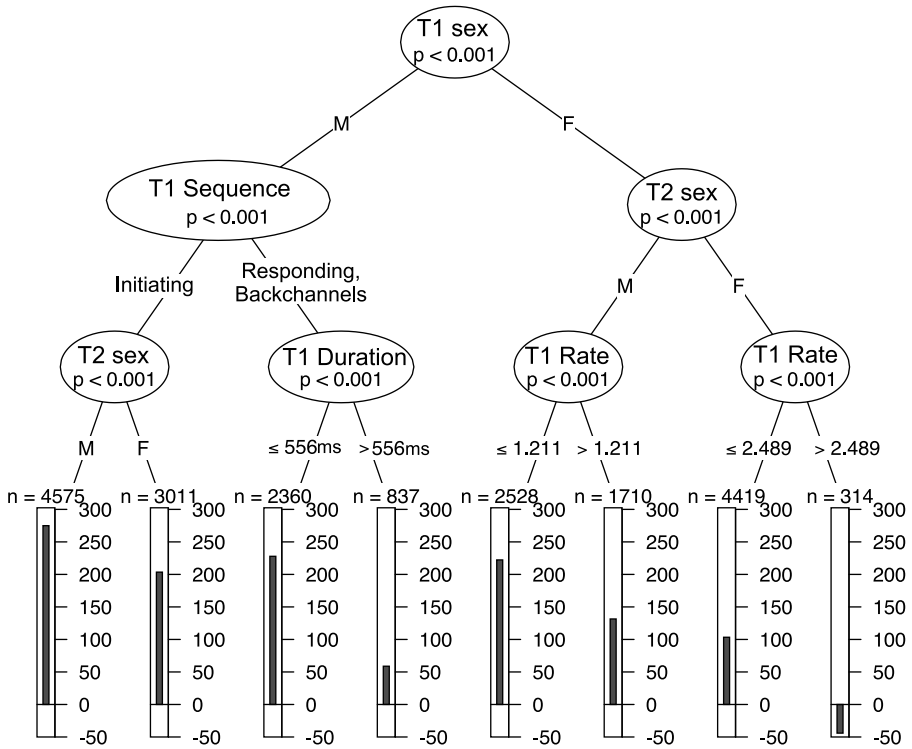


FIGURE 2 | A decision tree splitting data into gaps and overlaps by measures of sequence organization and processing.

tree. Once a large number of trees has been run, the relative importance of variables can be assessed.

We measure variable importance as the standard mean decrease in classification accuracy when a variable is permuted (see Breiman, 2001). For each tree in the forest, the prediction error (mean squared error) is calculated by comparing the true values of FTO to the values predicted by the tree. Taking the variable for which the measure is to be calculated, the assignment of each value of that variable to a case is randomly permuted and the prediction error is re-calculated. The difference between the two errors gives a measure of how influential the variable is for prediction of FTO. The difference in errors are calculated for all trees. The importance measure is then the mean of these differences normalized by the standard deviation of the differences.

The higher the importance value, the more influential the variable is in predicting the dependent variable.

For our purposes, random forests provide a way of assessing the relative importance of variables when the independent variables are highly correlated and when relationships between variables may be more complicated than simple linear patterns. Random forests have been used to look at various phenomena in linguistics (e.g., Bürki et al., 2011; Tagliamonte and Baayen, 2012; Plug and Carter, 2014; Sadat et al., 2014). Schneider (2014) analyzed the Switchboard corpus using binary decision trees and random forests to explore the distribution of hesitations in turns according to word co-occurrence frequency. Hesitations were less common between words that frequently co-occurred, supporting a "chunking" theory of language processing and production (e.g., Arnon and Snider, 2010; Bybee, 2010). However, this study did not consider the sequential organization of turns. We implement random forests using the functions *ctree* and *cforest* in the R package party (Hothorn et al., 2006a,b; Strobl et al., 2007, 2008).

Decision trees split data into subsets that can be modeled separately. That is, they try to find clusters of data that behave in similar ways. This is slightly different from linear regression which looks for linear relationships in the data as a whole. One prediction from the processing account might be that turns with low-frequency words will be responded to differently (slower) than other turns. Therefore, the tree would split the data into FTOs with high and low frequency T1s. A prediction from the sequence organization account might be that negative responses have higher FTOs, so the tree would split the data into FTOs before positive and negative T2 responses.

7. Results

The distribution of FTOs is shown in **Figure 3**. The mean FTO was 187 ms; the median was 168 ms; the standard deviation was 448 ms; the mode (calculated by gaussian kernels with the *density* function in R set to default parameters) was 169 ms. For comparison, in our Switchboard data, the median for polar questions followed by a response was 199 ms, and (Stivers et al., 2009) found that the median FTO for polar questions followed by a response was 200 ms.

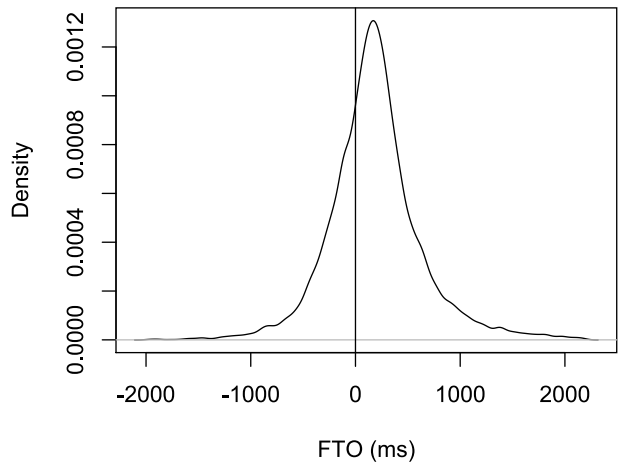


FIGURE 3 | The distribution of floor transfer offsets (the gap between two turns) for the Switchboard data.

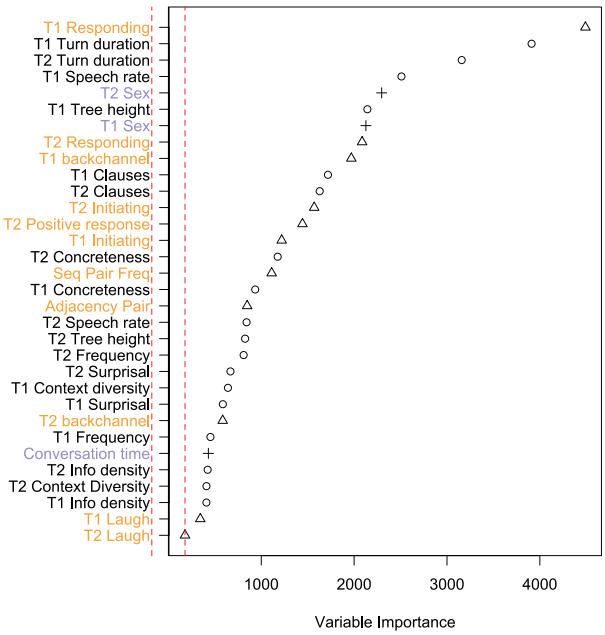


FIGURE 4 | Variable importance in a random forest analysis of floor transfer offset. The dotted red line shows the absolute smallest value, which can be used as a baseline for spurious effects. Measures of processing appear as circles (black labels) and measures of sequence organization appear as triangles (orange labels). Other measures appear as crosses (purple labels).

Many of the predictor variables are correlated with each other (three quarters of the variables were correlated with $p < 0.05$, see the Supplementary Materials), though there was only weak evidence for multicollinearity (maximum variable inflation factor = 3.9). The number of variables also makes the number of possible interactions very high. These two factors make simple linear regression analyses more difficult to interpret, but random forests is a robust to these concerns. Here we report various results relating to the random forests analyses.

A random forests model was run with 1000 trees and 3 variables in each tree (two runs of the model with different random stating seeds produced highly correlated variable importance measures, $r = 0.996$, $df = 30$, $p < 0.001$, suggesting that the results are robust, the results are also highly correlated when using 5 variables in each tree, see the Supplementary Materials 1). To give an impression of the fit of the model, a single tree was generated (like in the example above, but not limited to 3 levels). The predicted FTOs correlated with the actual FTOs with $r = 0.51$, meaning that the model accounts for about 30% of the variation. Another way of assessing the fit is to use the model to predict values for each FTO. When categorizing FTOs into gaps and overlaps, the model correctly categorizes 70% of cases.

In comparison, a simple linear model accounts for about 4% of the variance in FTO (see Supplementary Materials). This result is difficult to compare with linear models, since random forests work very differently (random forests are based on decision trees which divide data into sub-sets and fit each sub-set separately). Still, the difference between the two suggests that overall trends are weak, but there are more dependable patterns for certain types of transition.

Figure 4 shows the importance estimate for each variable, as calculated by the Random Forests analysis. This is an indicator of the relative importance of each variable in explaining the variation in FTO. The baseline for spurious variables is set as the absolute lowest importance measure. All variables have a positive importance value.

The top five most important variables are whether T1 includes a responding action, T1 duration, T2 duration, T1 speech rate and T1 sex. Measures of processing and sequence organization were not rated differently overall (mean importance for processing measures = 1300, mean importance for sequence organization measures = 1387, $t = -0.21$, $p = 0.83$; mean rank for processing measures = 17.8, mean rank for sequence organization measures = 16.1 $t = 0.47$, $p = 0.64$).

There was no large difference in the ranking of measures for T1 compared to measures for T2 ($t = -0.63$, $df = 26$, $p = 0.53$). For duration, speech rate and tree height the importance of the variable for T1 is greater than for T2, suggesting more weight on comprehension and planning. However, the opposite pattern holds for concreteness, frequency, and surprisal measures.

In the following sub-sections, we consider some of the most important variables, and comment on how they are related to FTO. The ranking of importance comes directly from the model results. However, the relationship with FTO is not easy to extract from the model, since a particular variable may be used to divide cases into sub-samples in very different ways. Therefore, when considering the relationship between a given variable and FTO, we explore the trends in the overall data.

7.1. Results for Measures of Processing

To give a sense of the overall trends for the processing measures, **Table 2** shows the simple, linear correlation between them and FTO (more straightforward descriptive results can be found in the Supplementary Materials). Most correlations are very

TABLE 2 | The Pearson correlation between processing measures and FTO for T1 and T2.

	T1	T2
Concreteness	0.028	−0.004
Mean frequency	−0.010	0.024
Speech rate	−0.091	−0.008
Information uniformity	−0.009	−0.004
Turn duration	0.043	0.025
Surprisal	−0.003	−0.012
Number of clauses	0.026	−0.019
Syntax tree height	0.065	0.012
Contextual diversity	−0.027	−0.014

weak, yet, as we show below, the random forests approach does find robust patterns. This suggests that the the relationship between measures of processing and FTO is complicated: overall tendencies are weak, but more dependable patterns can be found for certain types of transition.

7.1.1. Turn Duration and Rate

The variables ranked second and third most important are the durations of T1 and T2. Both T1 and T2 duration have similar relationships with FTO (see **Figure 5**). This relationship is non-linear: overall, longer turns occur with longer FTOs, but very short turns (less than 700 ms) are also followed by longer FTOs. The production and comprehension prediction was that longer turns would take longer to plan or comprehend, and therefore possibly start later. However, since T2 length is not linearly related to FTO, but the variable is ranked as highly important in the random forests analysis, this suggests that turn duration is being used as a proxy to distinguish different types of turn. Indeed, around three quarters of turns less than 700 ms are backchannels or agreements, while around three quarters of turns longer than 700 ms are statements, opinions, and questions. In line with this, the splits in the decision trees tend to divide data by turn duration at around 700 ms (e.g., see example decision tree in Supplementary Materials 2).

Speech rate of T1 is ranked as the 4th most important variable. On average, as T1 is spoken faster, the FTO becomes longer (this holds when excluding backchannels and short T1s). T1s spoken with rates in the fastest quartile lead to FTOs around 100 ms longer than those in the slowest quartile. The speech rate of T2 is ranked as much less important. There is no strong relationship between T2 rate and FTO.

7.1.2. Syntactic Complexity

T1 syntactic tree height is relatively important (ranked 6th most important out of 30), as is the number of clauses for T1 and T2 (ranked 9th and 10th). As the T1 increases in syntactic complexity, the FTO increases. Turns in the simplest quartile lead to FTOs 64 ms shorter than turns in the most complex quartile. There is no strong linear relationship between T2 syntactic complexity and FTO. The relative importance of the number of clauses in T2 may be attributed to the correlation with turn duration ($r = 0.65$, $t = 171$, $p < 0.00001$). Notice that here, as with speech rate, the processing factors only have significance in a particular sequential context, demonstrating how the two parameters, sequence organization and processing costs, are interwoven.

7.1.3. Concreteness

T2 concreteness is placed in the middle of the ranking. The prediction was that turns with more concrete words will lead to lower FTOs. However, the relationship with FTO is complicated. There is no overall linear relationship. There are interactions with turn duration so that there is a positive relationship for short T2s and a negative relationship for longer T2s. This could be explained in the following way: very short turns such as backchannels tend to have very low concreteness ratings. However, some short turns, such as answers to open questions have very concrete ratings (e.g., "How many kids do you have?," "Two"). When combined with utterance duration, then, the concreteness of T2 becomes a proxy for distinguishing response tokens (simple to project and plan) from question answers (more difficult to project and plan). Indeed, in a decision tree constructed with only T2 concreteness and T2 duration, T2 concreteness is used in a branch of the tree with short T2 turns and, within these turns, higher concreteness leads to longer average FTOs (positive relationship).

T2 concreteness seems to be more related to the absolute FTO, that is to how close the beginning of T2 is to the end of T1, ignoring whether it's a gap or overlap. There is an overall positive correlation between absolute FTO and T2 concreteness ($r = 0.13$), with the correlation being stronger as the length of T1's turn increases (for turns longer than 1000 ms, $r = 0.23$). That is the timing of turn transition is more tightly timed when T2 is less concrete (especially for longer T1s).

The relationship between T1 concreteness and FTO is more straightforward. T1s with low mean concreteness ratings are followed by short FTOs, while mid-range concreteness ratings have longer FTOs. However, T1s with high mean concreteness ratings have lower FTOs than mid-range turns.

7.2. Results for Measures of Sequence Organization

7.2.1. Initiating and Responding Actions

The most important factor in the whole random forests analysis is whether T1's (final) dialog act includes a responding action (e.g., an answer to a question). On average, FTOs are smaller when T1 includes a responding action (150 ms, compared to 202 ms in other cases, *post-hoc* $t = 7.9$, $p < 0.00001$). Whether T2 starts with a responding action, and whether T2 starts with an initiating action are also ranked as relatively important, and since they form the basis of sequence organization they are discussed together here. **Figure 6** shows the mean FTOs for different combinations of T1 and T2 sequence types.

The mean FTO when T1 initiates and T2 responds (e.g., a question in T1, followed by an answer in T2) is 200.7 ms. This kind of sequence forms the basis of adjacency pairs (see Section 4.1), and agrees very well with results for polar questions from Stivers et al. (2009).

The mean FTO is longer when T1 responds and T2 initiates (284.8 ms), a floor transition involving turns which do not form an adjacency pair. For example, in the extract below, B asks a question ("lots of little funny spots, huh?"), and A gives an answer ("Oh, yeah, yeah"). This is a well-formed adjacency pair. However, if we consider the answer as T1, the next turn T2 is a different question from B. These latter turns are not part of an adjacency pair, but belong to other sequences.

Conversation 3254, 0:19

(A and B are comparing a modern adaptation of the Adams Family with the original TV series, which includes a character called Thing)

```
    A: Uh, there were a few things different
       than the old series, but on the, on
       the whole, it was pretty similar.
       And, a lot of fun.
    B: Lots of little funny spots, huh?
T1  A: Oh, yeah, yeah. (Responding)
    FTO = + 614 ms
T2  B: Did they have Thing, and,
       (Initiating)
    A: Oh, yes, in fact, Thing has a big,
       much bigger role than he does in the
       series.
```

Another possible case is that in which floor transfers occur between two initiating actions. In such cases, the mean FTO was the longest (298 ms). In our data, these often involve cases of other-initiated repair (34% of all repair initiators occur in a transition where T1 and T2 include initiating actions; 40% of turn transitions where both T1 and T2 include initiating actions involve repair). The following example is a case of other initiated repair. B asks a question (initiating action), but A does not hear or understand, and initiates a repair sequence on the previous turn. B goes on to rephrase their question, and A resumes the main question-answer sequence:

Conversation 3232, 2:13

(A and B are discussing scholarships)

```
    B: -- it paid most of my tuition, and,
       um, a lot of the book costs and that
       kind of thing, so.
    A: Wow, that's great.
    B: Yeah, I really,
T1  A: Was it a Pell grant? / (Initiating)
    FTO = +494 ms
T2  B: I'm sorry, what did you say?
       (Initiating)
    A: What kind of grant was it?
    B: Well, it was called a B E O G,
       a Basic Equal Opportunity Grant
```

In line with our results, Kendrick (2015) finds that repair initiators are delayed compared to answers to questions.

Finally, we see that the shortest average FTOs are when both T1 and T2 involve responding actions. In this case, the mean FTO is shorter (157 ms). Many of these sequences involve T1 being a backchannel. Looking closer, we also find that many are

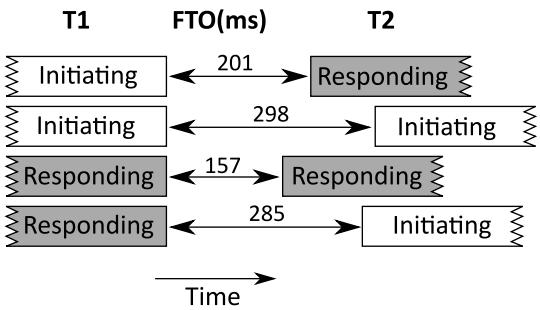


FIGURE 6 | Mean FTOs between turns with different kind of sequential actions (responding and initiating).

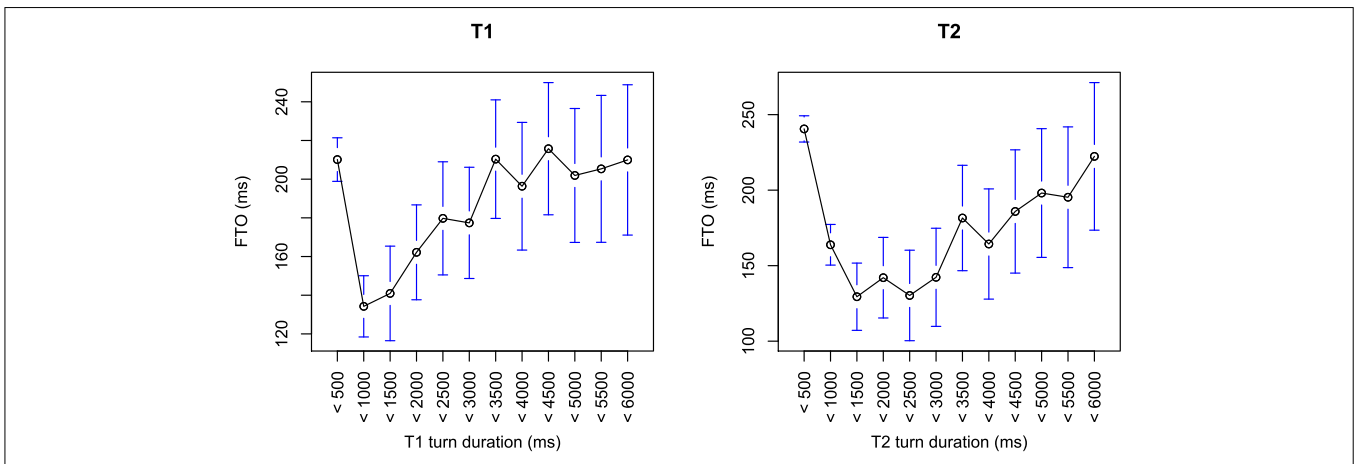


FIGURE 5 | The relationship between FTO and T1 and T2 duration. The data is grouped into 500 ms bins. Circles represent the mean of the bin, with bars showing the 95% confidence intervals.

part of sequences of assessment. The example below involves a news delivery sequence (e.g., Maynard, 1997). A announces some news and B delivers a news receipt ("wow"), after which there are several elaboration turns with assessments.

Conversation 3201 3:00

(A is talking about a recycling service)

```
    A: But, uh, they just go around to
       each, uh, door and pick it up.
    B: Wow, that's excellent.
T1  A: Yeah. (Responding)
    FTO = +137 ms
T2  B: That's good. (Responding)
```

Similarly, in the next example, A responds to B's statement with an assessment in T1, and then B produces a second assessment in T2.

Conversation 3526 2:44

(A and B are talking about cheap computers sold at a local warehouse)

```
    B: Yeah it and when it comes on the
       manufacturing floor it's about ten
       bucks
T1  A: I'll be darned (Responding)
    FTO = +206
T2  B: Yeah (Responding)
    A: Huh well
    B: Well i watched something on TV a
       couple of months ago by uh General
       ex uh Surgeon General Koop
```

7.2.2. Backchannels

T1 including a backchannel is ranked 9th out of 30 in terms of importance. Looking at the effect of this variable *post-hoc*, we observed that, when T1 is a backchannel, the FTO is around 38 ms lower on average than otherwise. This could occur if backchannels are regularly overlapped because they are not treated as real turns at talk (see SL & FT in Levinson et al., 2015). Indeed, the average FTO when T1 includes a backchannel is lower (with backchannel = 157 ms, without backchannel = 194 ms, *post-hoc* $t = 5.43$, $p < 0.0001$; 35% of T1 backchannels are overlapped by the next turn, compared to 28% of other cases). There is no big difference in means according to whether T2 is a backchannel (with backchannel = 195 ms, without backchannel = 183 ms, $post - hoct = -1.9$, $p = 0.06$) and this is reflected in it being ranked as relatively unimportant in the random forests results.

7.2.3. Positive Responses

Whether T2 provides a positive response is ranked in the middle of the distribution of importance. *Post-hoc* tests revealed that FTOs are 55 ms longer on average when T2 includes a negative response ($t = 2.38$, $df = 1348$, $p = 0.02$). This is in line with a delay for dispreferred responses, but the size of the effect is very small (the effect is weaker in a mixed effects model controlling for speaker identity and dialect), especially in comparison to the effect of T2 being a responding action vs. not.

7.3. Other Effects

The sex of the speakers is relatively important, with each male in the conversation adding around 70 ms on average to the FTO (similar differences are obtained from a mixed effects model controlling for speaker identity and speaker dialect).

The rest of the variables have weaker importance values, but some observations are worth making. Many processing variables are not highly ranked, especially measures of information and surprisal, but also frequency, which goes against the processing predictions.

FTOs are on average lower for transitions involving laughter (mean without laughter = 192 ms, mean with laughter in T1 or T2 or both = 112 ms, *post-hoc* $t = 5.4$, $p < 0.00001$). FTOs are shortest when there is invited laughter: when there is laugher in both T1 and T2, the average FTO is -142 ms (overlap), as predicted by the literature on laughter in conversation. However, the laugher variables are the lowest rated variable according to the random forests analysis. This could be due to the relatively small number of cases that include laughter (about 4% of cases).

8. Model Without Turn-Preserving Placeholders

As discussed in Section 7, the beginnings of some turns may be turn-preserving placeholders, hesitation markers such as "um" and "uh," that speakers use to "buffer" their response. This could obscure the demands on processing. To explore this, the same model was run, but calculating the FTO as the time from the end of T1 to the first non-turn-preserving placeholder in T2. The full results are available in the Supplementary Materials. The importance estimates in this model were weakly correlated with the main model importance estimates reported in the section above ($r = 0.597$, $df = 30$, $p = 0.0003$; rank correlation = 0.73). The prediction from processing is that the processing variables would be ranked as more important in this case, since placeholders gives responders time to plan.

The main difference in this model is that T2 turn duration has increased in importance. That is, the length of T2's turn is a better predictor of gap duration when turn-preserving placeholders are ignored. This could be evidence that speakers are "buffering" turns which require more planning. Overall, however, the processing measures do not become more important on average. Also, measures for T2 did not increase in relative importance compared to measures for T1.

Therefore, while there is some evidence that turn-preserving placeholders do buffer planning, the importance of sequential organization variables in explaining FTO cannot be easily attributed to this effect.

9. Discussion

This paper has examined explanations for the timing of turn taking taken either from hypotheses about cognitive processing or from those originating from sequence organization. Neither processing nor sequence organization dominated as important measures. Basic sequence organization measures such as the

sequential status of turns were informative, as were measures of turn duration, speech rate, and syntactic complexity. Perhaps unexpectedly, measures of frequency and surprisal were ranked as much less important, even though they are known to affect processing and production of language in laboratory conditions. This suggests that, in real conversation, these effects often only apply in specific sequential positions, e.g., in T1 or T2, or when T1 is initiating, showing that the two kinds of account are interwoven.

One question is the suitability of the measures used. The measures of processing, for example, are not direct measures of cognitive activity, but properties of utterances that are known to correlate with processing. Having said this, the sequence organization measures are also very coarse, suggesting that this is not biasing the comparison between the two domains. Obviously, more direct on-line measurement of processing during conversation would be ideal (e.g., Holler and Kendrick, 2015). Experimental control and ecological validity are difficult to balance, but this study suggests that such an approach is warranted in the future.

In some cases, there was a difference in the predictions for processing and sequence organization accounts. For speech rate, we found that faster speech is responded to with longer FTOs. This fits with a processing account rather than a straightforward "rhythm" account, which would predict that a faster beat would lead to a faster response. Although speech rate is not the same as rhythm, and a more suitable analysis would be to code the dataset for rhythm, we still find little support for the idea that responders generally respond on the first beat, all other things being equal.

The random forests analysis explained a reasonable amount of the variance in FTOs. While it's clear that the timing of turn taking is a noisy process, the analysis suggests that there are systematic principles. The relatively good performance of the random forests analysis compared with the linear analysis also suggests that the principles are context-sensitive, rather than applying across the board. For example, certain processing effects being only evident in certain sequential contexts.

Indexical information such as the sex of each speaker was ranked as relatively important. However, these differences may not be linked inherently to sex (e.g., through processing differences), but may reflect differences in socio-cultural norms or simply individual differences.

10. Conclusion

We began with the observation that communication systems are imaginable, the marmoset system a putative case, in which issues of cognitive load and sequence organization play little role in influencing temporal patterns of behavior. Human communication contrasts on both dimensions, because of the formidable choice of alternatives faced by a speaker and the consequent unpredictability faced by a responder on the one hand, and on the other because sequences of ordered turns map structure onto the sheer fact that T1 is followed by T2.

By using a large coded corpus we have been able to track the importance of a set of different measures of each dimension. We conclude that the temporal patterns of dialogue cannot be accounted for by either cognitive or sequence organization factors alone. The two are interwoven with indexical factors in such a way that, for example, the sex of a speaker in a particular initiating sequence type creates an environment where cognitive load plays a particularly strong role in influencing the speed of response. This suggests that an explanation of the timing of turn taking will involve a combination of insights from both cognitive processing and sequential organization.

The ways in which factors load only in specific ecological niches make standard regression techniques inapplicable. Here the method employed, random forests, comes into its own, allowing the factor loading to be discerned in specific ecological niche formed by indexical factors, processing factors and sequence factors, as illustrated in the tree in **Figure 2**. The kinds of binary decision trees produced in this paper make predictions that could also be tested experimentally. There is also the possibility of using real conversational data from the Switchboard corpus as stimuli material to create a cycle of qualitative analysis and quantitative testing (e.g. Kendrick and Torreira, 2015).

This study has not exhausted the information in the Switchboard corpus. It would be possible to analyze further the different distinctions in dialog actions, such as the distinction between polar questions and content questions (initial analyses suggest that polar questions are responded to up to 100 ms sooner than content questions). There is also more indexical information such as age and dialect. Speakers were assigned topics of conversation and these may also have stimulated speakers to different degrees, which could affect average FTOs. There are also a range of phonetic information and semantic factors that could be explored. This study thus has distinct limitations. Extending the analysis to other corpora and other languages will however require large amounts of transcribed speech data, matched with processing information such as frequency and surprisal for many languages.

10.1. Human Search and Animal Research

All data collected from individuals were from sources where informed consent had been provided.

10.2. Data Sharing

Source data is available to download online (see the various references in the main text). The analysis software is also available: pympi (Lubbers and Torreira, 2014); ELAN, developed by the Max Planck Institute for Psycholinguistics (Wittenburg et al., 2006) http://tla.mpi.nl/tools/tla-tools/elan/.

Author Contributions

Extracted and prepared the data: FT; Calculated FTO: FT; analyzed the data: SR, FT; Wrote the manuscript: SR, FT, SL.

Funding

SR is supported by the Interactional Foundations of Language project within the Language and

Cognition Department at the Max Planck Institute for Psycholinguistics.

Acknowledgements

Thanks to Mart Lubbers for technical support. This research was conducted within, and partially financed by, the framework of ERC Advanced Grant No. 269484 INTERACT to SL. We thank the Max Planck Society for additional support.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00509/abstract

References

Arnon, I., and Snider, N. (2010). More than words: frequency effects for multi-word phrases. *J. Mem. Lang.* 62, 67–82. doi: 10.1016/j.jml.2009.09.005

Atkinson, J. M., and Heritage, J. (1984). "Preference organization," in *Structures of Social Action: Studies in Conversation Analysis*, ed J. M. Atkinson (Cambridge: Cambridge University Press), 53–56. doi: 10.1017/CBO9780511665868

Bögels, S., and Torreira, F. (in press). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *J. Phonetics*.

Bürki, A., Alario, F. X., and Frauenfelder, U. H. (2011). Lexical representation of phonological variants: evidence from pseudohomophone effects in different regiolects. *J. Mem. Lang.* 64, 424–442. doi: 10.1016/j.jml.2011.01.002

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The english lexicon project. *Behav. Res. Methods* 39, 445–459. doi: 10.3758/BF03193014

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Brysbaert, M., and New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A., and Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behav. Res. Methods* 46, 904–911. doi: 10.3758/s13428-013-0403-5

Bybee, J. L. (2010). *Language, Usage and Cognition*, Vol. 98. Cambridge: Cambridge University Press.

Calhoun, S., Carletta, J., Brenier, J. M., Mayo, N., Jurafsky, D., Steedman, M., et al. (2010). The NXT-format switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Lang. Resour. Eval.* 44, 387–419. doi: 10.1007/s10579-010-9120-1

Clark, H. H., and Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition* 84, 73–111. doi: 10.1016/S0010-0277(02)00017-3

Clark, H. H. (1976). *Semantics and Comprehension, Janua Linguarum. Series Minor 187*. Boston: DE GRUYTER MOUTON.

Clayman, S. E., Heritage, J., Elliott, M. N., and McDonald, L. L. (2007). When does the watchdog bark? Conditions of aggressive questioning in presidential news conferences. *Am. Soc. Rev.* 72, 23–41. doi: 10.1177/000312240707200102

Clayman, S. E. (2002). Sequence and solidarity. *Adv. Group Process.* 19, 229–253. doi: 10.1016/S0882-6145(02)19009-6

Couper-Kuhlen, E. (1993). *English Speech Rhythm: Form and Function in Everyday Verbal Interaction,* Vol. 25. Philadelphia, PA: John Benjamins Publishing.

De Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

Deshmukh, N., Ganapathiraju, A., Gleeson, A., Hamaker, J., and Picone, J. (1998). "Resegmentation of switchboard," in *ICSLP* (Syndey).

Drew, P. (2013). "Turn design," in *The Handbook of Conversation Analysis,* eds J. Sidnell and T. Stivers (Hoboken, NJ: Wiley Online Library), 131–149.

Ford, C. E., and Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. *Stud. Interact. Sociolinguist.* 13, 134–184.

Gardner, R. (2001). *When Listeners Talk: Response Tokens and Listener Stance*, Vol. 92. Philadelphia, PA: John Benjamins Publishing.

Gisladottir, R. S., Chwilla, D. J., Schriefers, H., and Levinson, S. C. (2012). "Speech act recognition in conversation: experimental evidence," in *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (Sapporo), 1596–1601.

Glenn, P. J. (1989). Initiating shared laughter in multi-party conversations. *West. J. Commun.* 53, 127–149.

Glenn, P. J. (2003). *Laughter in Interaction*. Cambridge: Cambridge University Press.

Godfrey, J. J., Holliman, E. C., and McDaniel, J. (1992). "Switchboard: telephone speech corpus for research and development," in *1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992. ICASSP-92,* Vol. 1, (San Francisco, CA), 517–520.

Haakana, M. (2002). Laughter in medical interaction: from quantification to analysis, and back. *J. Sociolinguist.* 6, 207–235. doi: 10.1111/1467-9481.00185

Hanley, J. R., Hunt, R. P., Steed, D. A., and Jackman, S. (2013). Concreteness and word production. *Mem. Cogn.* 41, 365–377. doi: 10.3758/s13421-012-0266-5

Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *J. Phonet.* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002

Heritage, J. (1984). "A change-of-state token and aspects of its sequential placement," in *Structures of Social Action: Studies in Conversation Analysis*, ed J. M. Atkinson (Cambridge: Cambridge University Press), 299–345. doi: 10.1017/CBO9780511665868

Holler, J., and Kendrick, K. H. (2015). Unaddressed participants gaze in multi-person interaction: optimizing recipiency. *Front. Psychol.* 6:98. doi: 10.3389/fpsyg.2015.00098

Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., and Van Der Laan, M. J. (2006a). Survival ensembles. *Biostatistics* 7, 355–373. doi: 10.1093/biostatistics/kxj011

Hothorn, T., Hornik, K., and Zeileis, A. (2006b). Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat.* 15, 651–674. doi: 10.1198/106186006X133933

Jefferson, G. (1974). *Notes on the Sequential Organization of Laughter in Conversation: Onset Sensitivity in Invitations to Laugh*. Mexico City: American Anthropological Association Convention.

Jefferson, G. (1984). "On the organization of laughter in talk about troubles," in *Structures of Social Action: Studies in Conversation Analysis*, Vol. 346369, eds J. M. Atkinson and J. C. Heritage (Cambridge: Cambridge University Press), 346–369.

Jefferson, G. (1986). Notes on latencyin overlap onset. *Hum. Stud.* 9, 153–183.

Jurafsky, D., Shriberg, E., and Biasca, D. (1997). *Switchboard SWBD-DAMSL Shallowdiscourse-Function Annotation Coders Manual*. Institute of Cognitive Science Technical Report, Boulder, 97–102.

Kemper, S., and Rash, S. (1988). Speech and writing across the life-span. *Pract. Aspects Mem. Curr. Res. Issues* 2, 107–112.

Kemper, S., Kynette, D., Rash, S., O'Brien, K., and Sprott, R. (1989). Life-span changes to adults' language: effects of memory and genre. *Appl. Psycholinguist.* 10, 49–66.

Kendrick, K. H., and Torreira, F. (2015). The timing and construction of preference: a quantitative study. *Discourse Process* 52, 255–289. doi: 10.1080/0163853X.2014.955997

Kendrick, K. H. (2015). The intersection of turn-taking and repair: the timing of other-initiations of repair in conversation. *Front. Psychol.* 6:250. doi: 10.3389/fpsyg.2015.00250

Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. (2006). "On the human interaction engine," in *Roots of Human Sociality: Culture, Cognition and Human Interaction*, eds N. Enfield and S. Levinson (Oxford: Berg), 39–69.

Levinson, S. C. (2013a). "Action formation and ascription," in *The Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Malden, MA: Wiley-Blackwell), 103–130.

Levinson, S. C. (2013b). Recursion in pragmatics. *Language* 89, 149–162. doi: 10.1353/lan.2013.0005

Levinson, S. C., Holler, J., Kendrick, K. H., and Casillas, M. (2015). Turn-taking in human communicative interaction. *Front. Psychol. Res. Top.* Available online at: http://journal.frontiersin.org/researchtopic/turn-taking-in-human-communicative-interaction-2845

Liberman, M. (2014), *Um, there's timing information in switchboard?* Available online at: http://languagelog.ldc.upenn.edu/nll/?p=14991

Local, J. (1996). Conversational phonetics: some aspects of news receipts in everyday talk. *Stud. Interact. Sociolinguist.* 12, 177–230.

Lubbers, M., and Torreira, F. (2014). *Pympi Python Library Version 1.1.* Available online at: https://github.com/dopefishh/pympi

Magyari, L., and De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

Marcus, M. P., Santorini, B., Marcinkiewicz, M. A., and Taylor, A. (1999). *Treebank-3.* (Linguistic Data Consortium Catalog (LDC) Catalog LDC99T42).

Maynard, D. W. (1997). The news delivery sequence: bad news and good news in conversational interaction. *Res. Lang. Soc. Interact.* 30, 93–130.

Piantadosi, S., Tily, H., and Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3526–3529. doi: 10.1073/pnas.1012551108

Plug, L., and Carter, P. (2014). Timing and tempo in spontaneous phonological error repair. *J. Phonet.* 45, 52–63. doi: 10.1016/j.wocn.2014.03.007

Potts, C. (2011). *The Switchboard Dialog Act Corpus.* Available online at: http://compprag.christopherpotts.net/swda.html

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 40, 696–735.

Sadat, J., Martin, C. D., Costa, A., and Alario, F. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cogn. Psychol.* 68, 33–58. doi: 10.1016/j.cogpsych.2013.10.001

Schegloff, E. A. (2007). *Sequence Organization in Interaction,* Vol. 1. *A Primer in Conversation Analysis,* Vol. 1. Cambridge: Cambridge University Press.

Schneider, U. (2014). "Cart trees and random forests in linguistics," in *Hochleistungsrechnen in Baden-Wuerttemberg-Ausgewaehlte Aktivitaeten im bwGRiD 2012: Beitraege zu Anwenderprojekten und Infrastruktur im bwGRiD im Jahr 2012*, eds J. C. Schulz and S. Hermann (Karlsruhe: KIT Scientific Publishing), 67.

Schwanenflugel, P. J., Harnishfeger, K. K., and Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *J. Mem. Lang.* 27, 499–520.

Stivers, T., and Enfield, N. J. (2010). A coding scheme for question–response sequences in conversation. *J. Pragmat.* 42, 2620–2626. doi: 10.1016/j.pragma.2010.04.002

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Stolt, M. (2008). *The Many Faces of Overlap: Non-Competitive Overlap in a Conversation between Finnish and British Speakers of English.* Ph.D. thesis, University of Jyväskylä.

Strömbergsson, S., Hjalmarsson, A., Edlund, J., and House, D. (2013). "Timing responses to questions in dialogue," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2013* (Lyon: International Speech and Communication Association), 2584–2588.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures. *BMC Bioinformatics* 8, 1–10. doi: 10.1186/1471-2105-8-25

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics* 9:307. doi: 10.1186/1471-2105-9-307

Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* 14, 323. doi: 10.1037/a0016973

Tagliamonte, S. A., and Baayen, R. H. (2012). Models, forests, and trees of york english: was/were variation as a case study for statistical practice. *Lang. Variat. Change* 24, 135–178. doi: 10.1017/S0954394512000129

Takahashi, D. Y., Narayanan, D. Z., and Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* 23, 2162–2168. doi: 10.1016/j.cub.2013.09.005

Torreira, F., Bögels, S., and Levinson, S. C. (2015). Breathing for answering: the time course of response planning in conversation. *Front. Psychol.* 6:284. doi: 10.3389/fpsyg.2015.00284

Walker, I., and Hulme, C. (1999). Concrete words are easier to recall than abstract words: evidence for a semantic contribution to short-term serial recall. *J. Exp. Psychol.* 25, 1256.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 91, 1707–1717.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "Elan: a professional framework for multimodality research," in *Proceedings of LREC, 2006* (Genoa).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 March 2015
doi: 10.3389/fpsyg.2015.00284

Breathing for answering: the time course of response planning in conversation

Francisco Torreira[1]*, *Sara Bögels*[1] and *Stephen C. Levinson*[1,2]

[1] Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

We investigate the timing of pre-answer inbreaths in order to shed light on the time course of response planning and execution in conversational turn-taking. Using acoustic and inductive plethysmography recordings of seven dyadic conversations in Dutch, we show that pre-answer inbreaths in conversation typically begin briefly after the end of questions. We also show that the presence of a pre-answer inbreath usually co-occurs with substantially delayed answers, with a modal latency of 576 vs. 100 ms for answers not preceded by an inbreath. Based on previously reported minimal latencies for internal intercostal activation and the production of speech sounds, we propose that vocal responses, either in the form of a pre-utterance inbreath or of speech proper when an inbreath is not produced, are typically launched in reaction to information present in the last portion of the interlocutor's turn. We also show that short responses are usually made on residual breath, while longer responses are more often preceded by an inbreath. This relation of inbreaths to answer length suggests that by the time an inbreath is launched, typically during the last few hundred milliseconds of the question, the length of the answer is often prepared to some extent. Together, our findings are consistent with a two-stage model of response planning in conversational turn-taking: early planning of content often carried out in overlap with the incoming turn, and late launching of articulation based on the identification of turn-final cues.

Keywords: breathing, turn-taking, conversation, language planning, language production, speech planning, speech production, turn projection

OPEN ACCESS

Edited by:
F-Xavier Alario, CNRS and Aix-Marseille Université, France

Reviewed by:
Amélie Rochet-Capellan, CNRS, France
Martin Corley, University of Edinburgh, UK

***Correspondence:**
Francisco Torreira, Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, Netherlands
francisco.torreira@mpi.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *20 November 2014*
Accepted: *26 February 2015*
Published: *12 March 2015*

Citation:
Torreira F, Bögels S and Levinson SC (2015) Breathing for answering: the time course of response planning in conversation. Front. Psychol. 6:284. doi: 10.3389/fpsyg.2015.00284

Introduction

Conversation is the core ecological niche for language—it is where language is learnt and most heavily used. Conversation is characterized by the rapid alternation of speakers, who each take mostly small turns at talk, generally avoid vocal overlap, and minimize the gap between turns (Sacks et al., 1974). This behavior appears to be, with minor wrinkles, universal in character (Stivers et al., 2009). Despite the universality and stability of this behavior, conversational turn-taking has figured little in theories about language processing, even though it poses a fundamental puzzle for them. Reported average inter-turn gap durations in the literature fall between 0 and 300 ms (e.g., De Ruiter et al., 2006; Stivers et al., 2009; Heldner and Edlund, 2010), but the latencies in language planning for production are much longer: it takes 600–1200 ms (depending on word frequency) to begin labeling a picture of an object from the moment it becomes visible (Levelt et al., 1999; Indefrey and Levelt, 2004), and it takes 1500 ms to begin producing a simple sentence describing an

action picture (Griffin and Bock, 2000). Even allowing for contextual priming and facilitation, these latencies are substantial. This would seem to rule out the idea that participants simply respond to turn ends: the fastest human reaction times are of the order of 100–200 ms and the minimal latency reported for a pre-rehearsed syllable is 210 ms (Fry, 1975). Moreover, the speech signal has many brief moments of silence related to the ongoing linguistic signal (e.g., stop consonant closures), often lasting a similar duration to inter-turn gaps. So one could not recognize a silent gap as a gap before approximately 100–200 ms which, combined with minimal reaction time latency, would yield a conversational gap of 300–400 ms. These figures leave no time for the 500–1200 ms planning latencies of speech production discussed above, so the paradox of quick responses using a slow production system persists.

A plausible solution to the paradox is that, as foreseen in Sacks et al. (1974), responders often predict the content of the incoming turn, which allows them to begin planning a relevant response in advance of the turn end. The question still remains how listeners know *when* to articulate their response without causing unwanted overlap or long silent gaps (which may be semiotically loaded; cf. Kendrick and Torreira, 2015). Two possibilities can be envisaged. First, as proposed in a long tradition of observational studies (e.g., Duncan, 1972; Wells and MacFarlane, 1998; Caspers, 2003; Local and Walker, 2012), responders may launch articulation upon identifying turn-final cues (e.g., phrase-final melodic patterns, final lengthening, specific bodily gestures) occurring in the last syllables of their interlocutor's turn. This strategy could produce short gaps of 100 or 200 ms, only if at least the initial linguistic material of the responder's turn is ready to be articulated by the time the interlocutor's turn comes to an end.

An alternative option is that responders not only predict the content of incoming turns well in advance, but also estimate their timing on the basis of this prediction, and adjust the time course of their production planning based on such temporal estimation. In support of this view, for instance, De Ruiter et al. (2006) doubt that turn-final cues such as phrase-final intonation patterns are of any use for purposes of turn-taking, since they may occur too late in the turn to allow the listener to anticipate its end. Along the same lines, Magyari and de Ruiter (2012) state that it is very plausible that listeners know more than half a second in advance that a turn is going to end. Based on the results of a gating experiment, they propose that listeners make predictions in advance about which words and how many words will follow a partially heard turn, and that they use this prediction in order to estimate the remaining duration of that turn.

In this article, we explore the time course of response planning in conversation by focusing on an ancillary source of information about language production neglected so far in psycholinguistic discussions of turn-taking, namely, breathing. Several studies have identified a relationship between breathing behavior and utterance duration, indicating that breathing can be informative about the scope of language planning (e.g., Winkworth et al., 1995; Whalen and Kinsella-Shaw, 1997; Fuchs et al., 2013; Rochet-Capellan and Fuchs, 2013). Fuchs et al. (2013) investigated several speech planning parameters, including inhalation depth and inhalation duration, using read materials varying in length and syntactic complexity, and found that inhalation depth and duration were positively correlated with utterance length. Using a corpus of spontaneous conversation, Rochet-Capellan and Fuchs (2013) also observed positive correlations between utterance length and inbreath depth and duration. Given that pre-utterance inbreaths are indicative of the length of upcoming utterances, and that in spontaneous conversation they usually take over half a second to complete (McFarland, 2001), their timing with respect to an interlocutor's turn end in a turn-taking situation may offer interesting insights into the time course of the response planning process. If listeners estimate the timing of turn ends half a second or more in advance of the turn end to time their own response, for instance by predicting the final words of a turn and their duration (cf. Magyari and de Ruiter, 2012), we should observe that they often inhale well in advance of turn ends so that their response can be produced at the right moment (much like singers and wind instrument players do in advance of their musical entries as specified in the score). If, on the other hand, responders typically determine the position of turn ends on the basis of turn-final information, we should observe that pre-utterance inbreaths tend to be taken close to the end of the interlocutor's turn, and that answers preceded by an inbreath are substantially delayed compared to answers produced on residual breath. These two alternative mechanisms, early anticipation vs. local detection of turn ends, are presented schematically in **Figure 1**.

A relevant issue concerning the design of this study is whether we should investigate all conversational turns in a corpus across the board, or whether we should focus instead on a specific, more controlled, conversational context. While the first approach has the advantage of potentially producing more generalizable results, it may prevent us from uncovering relevant trends in the data due to uncontrolled sources of variability. For instance, contexts in which floor changes are optional involve the complexity of deciding who will be the next speaker, which will affect the timing of the next turn in unpredictable ways. To overcome such difficulties, we have decided to focus on inbreaths taken before

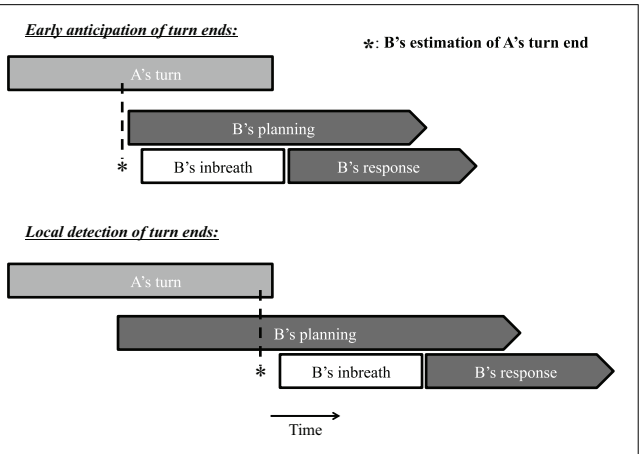


FIGURE 1 | Schematic representation of two possible response production mechanism involving a pre-utterance inbreath (see text for details).

answers in question-answer sequences, in which a conversational response is explicitly requested by the current speaker. In this sense, question-answer sequences provide an optimal conversational context in which to begin studying the time course of language planning during conversational turn-taking (cf. Stivers et al., 2009). Moreover, question-answer sequences can be identified in a reasonably objective way on the basis of the morphosyntactic and intonational marking of questions, and of the recognizability of following turns as relevant answers (cf. Stivers and Enfield, 2010). Finally, and importantly, question-answer sequences are one of the most common action sequences in spontaneous conversation, and can therefore provide a sufficient number of observations in a medium-sized corpus such as the one used in this study (see Section Material and Data Extraction).

Because we intend to use pre-utterance inbreaths as indicators of the time course and scope of language planning, we will first assess whether breathing behavior is related to utterance length in our data, as found in previous studies. We will also need to control for the role of other communicative factors that may be at play in conversational data. It is possible that, in conversation, pre-utterance inbreaths function as meaningful elements tied to the upcoming utterance, rather than as mere preparatory phases of upcoming utterances (cf. Schegloff, 2006). Kendrick and Torreira (2015) studied the timing and construction of preferred and dispreferred responses to invitations, offers, and requests (i.e., acceptances vs. rejections) in a corpus of telephone calls in English, and found that dispreferred responses tend to be preceded by an audible inbreath more often than preferred responses. It is therefore possible that in dispreferred responses, speakers want to make their pre-utterance inbreaths salient for the listener to indicate the preference status of their responses in advance, and that, for this reason, they avoid taking them in complete overlap with the interlocutor's turn. Rochet-Capellan and Fuchs (2013), also using spontaneous conversational materials, observed that utterances containing vocalized hesitations were preceded by deeper inhalations. An anonymous reviewer notes that, because of this, it is possible that pre-utterance inbreaths are also produced by speakers as indicators of hesitations and disfluencies in their upcoming utterances, and that this may constitute another reason for answerers to avoid taking inbreaths in overlap with the interlocutor's turn. In order to better assess the relationship between breathing behavior and language planning in our statistical analyses, we will take into account the preference status of the response, and the presence of disfluencies in the response.

The following section presents a description of an audiovisual corpus of spontaneous conversation in Dutch including inductive plethysmography recordings of respiratory activity, the extraction and coding of question-answer sequences from this corpus, and the measurement scheme applied to the data. Section Results then presents several statistical analyses aimed at answering the research questions discussed above, namely, (a) whether the characteristics of pre-utterance inbreaths in spontaneous conversation are related to the scope of language planning, (b) whether responses preceded by an inbreath are delayed with respect to end of the interlocutor's turn compared to responses spoken on residual breath, and (c) what the most common timing of pre-utterance inbreaths is relative to the end of the interlocutor's turn. In section Discussion, we review and interpret our findings, and sketch a turn-taking response production mechanism accounting for both the most common trends in the data and previously reported estimates of language processing latencies.

Materials and Methods

Material and Data Extraction

The corpus collection procedure and its use for research purposes were approved by the Ethics Committee Faculty of Social Sciences of the Radboud University Nijmegen. The corpus collection took place in a sound-attenuated room at the Max Planck Institute for Psycholinguistics. We recorded seven dyadic conversations between Dutch male friends, all of them university students except one participant (a research assistant). The reason for only recording males is that inductive plethysmography measurements are obtained more reliably from male participants than from female participants. Each recording had a duration of around 45 min, for an approximate total of 6 h and 15 min of dyadic conversation. Participants were briefly instructed to entertain a conversation with their dyad partner while sitting on chairs placed 1.5–2 m from each other, and oriented toward each other at an angle of 120 degrees. Each participant took part in the recordings only once.

The recording equipment consisted of a high-definition camera placed in front of the speakers, Shure SM10A head-mounted microphones, and an InductotraceTM inductive plethysmography system. Each participant wore an Inductotrace band attached around his chest at the level of the axilla, each connected to one of the two channels of the Inductotrace unit, and a head mounted-microphone coupled to an amplifier. The speech and breathing signals were recorded simultaneously at a sampling frequency of 48 kHz via an A/D converter connected to a computer. The breathing signals exhibited an upward drift starting approximately 10 min into the recording. Such a drift has not been reported in previous studies using the Inductotrace system, perhaps because their recordings were much shorter than ours. In order to correct this drift, we approximated the signals with third-order polynomials using the polyfit Matlab function, and extracted their residuals. Finally, we smoothed the signals by downsampling them by a factor of 1000.

Coding and Measures

Data Extraction

Using Elan software (Wittenburg et al., 2006), we extracted and annotated all question and answer sequences in the data, excluding those that exhibited laughter or coughing by a participant. Wh-questions were identified on the basis of the presence of interrogative pronouns or adverbs in the utterance. Polar questions were identified on the basis of their syntactic properties (i.e., subject-verb inversion) or final intonation contour (i.e., low-rising, high-rising, or rising-falling-rising). Question and answer sequences were first identified by an assistant unaware of the purposes of the study. The first two authors then checked whether the cases identified by the assistant complied with the criteria mentioned above and only retained those that did ($n = 171$). Each dyad contributed between 15 and 30 question-answer sequences

(mean = 21.6) to the dataset. Each speaker contributed between 4 and 26 answers to the dataset (mean = 12.9, *SD* = 5.9). The first author then marked the beginning and end of each question and answer. At the phonetic level, the beginning and end of answers and questions were located with reference to acoustic events in the signal attributable to either a lexical item or a particle (e.g., *uhm*, *uh*). Mouth noises, clicks and breathing noises were therefore not treated as part of the questions and answers. The beginning of the question was located with reference to syntactic structure (e.g., wh-words). The end of the answer was placed at the first pause that coincided with points of completion both at the syntactic and intonational levels. All answers therefore consisted of at least one syntactically and intonationally coherent phrase. We also coded the preference status of all answers in our data. We coded as preferred responses all responses to polar questions that matched them in polarity (e.g., *yes* answers in the case of polar affirmative questions), and answers to wh-questions that provided the requested information. Dispreferred responses included all other types of responses (e.g., negative answers to polar affirmative questions, responses to wh-questions in which the responder acknowledged not knowing the relevant answer).

We then displayed the breathing signals aligned with the audio signals in ELAN software. For each question-answer sequence, we identified inbreaths (i.e., rising trajectories of the breathing signals) only if they started after the beginning of the question and before the answer ($n = 91$; 53.2%), since inbreaths that started before the question could not have been produced in response to it. The shape of answerers' breathing signals in the considered interval showed considerable variation, and, in this respect, contrast with the breathing patterns described in studies based on highly-controlled speech. The signals could be flat with a final inbreath, but also falling or rising (i.e., indicating exhalation or inhalation), or exhibit a mixture of the preceding types (e.g., initially falling or rising, then flat, and then rising in a final inbreath). Moreover, it is probable that a number of the inbreaths that fell in the considered time interval were not primarily designed for speech. They could instead have been part of initially vital or partly vital breathing cycles that happened to occur in overlap with the question and preceding the answer. Although prototypical vital and speech breathing cycles differ very clearly under highly-controlled conditions (vital cycles are said to be more symmetrical than speech cycles, i.e., with more equal inhalation and exhalation phases; McFarland, 2001), many of the breathing cycles in our spontaneous speech data had shapes that could not be straightforwardly attributed to speech preparation or vital breathing mechanisms. Given the rapid alternation of turns of uncertain length in conversation, speakers may use different strategies to preserve sufficient lung air for speaking: For example, they may take precautions to breathe early, they may halt exhalation, or they may fall back on interruption of their production to breathe midway (cf. Bailly et al., 2013, for an illustration of different types of breathing behavior in collaborative reading). Because such strategies could not always be identified in a straightforward way, we decided not to classify the inbreaths in a qualitative way. Instead, we looked for meaningful quantitative trends in the data, while keeping in mind that different kinds of breathing behaviors were present in it.

Inbreath Annotation

The onset and offset of each annotated inbreath was marked at the signal minimum and maximum by an assistant unaware of the purpose of the study. In some cases where there was a low plateau, the onset was located at the "elbow" located at the end of the plateau rather than at the absolute minimum. We also measured the amplitude of each inbreath, and later converted this measure to speaker-normalized *z* scores for statistical purposes (note that, since we were not particularly interested in absolute kinematic values, we did not calibrate the Inductotrace instruments). Since the amplitude values in our data are approximately normally distributed, the normalized amplitude range for each speaker should roughly extend from −2 to 2 (excluding outliers). **Figure 2** illustrates our measurement scheme. From the initial timing measurements, we computed the time alignment of the beginning of the answerer's inbreath relative to the end of the question (inbreath latency from now on), and also to the start of the answer. Finally, we computed the duration of the answer, and its latency relative to the end of the question (answer latency).

Statistical Procedure

In order to investigate statistical relationships between variables of interest, we fit mixed-effects regression models using the lme4 R package (Bates et al., 2014), and perform model comparisons using the anova() function in R (R Core Team, 2014). We compare null models (including only a fixed intercept and random intercepts for each speaker), reduced models (also including any relevant covariates that may explain part of the variability of the response variable, such as the preference status of the response and the presence of disfluencies), and a full model with an added fixed predictor term for the main independent variable of interest in the analysis (e.g., answer duration in the analysis of the occurrence of pre-utterance inbreaths). In cases in which a reduced model does not improve the fit of the null model ($\alpha = 0.05$), we compare the full model directly to the null model. In all models, we include interactions between the random factor speaker and any fixed predictors only if the interaction is statistically significant in a separate model comparison. Notice, however, that none of them affected the coefficients of the other factors in the model in a major way. For this reason, and for the sake of simplicity, we do not discuss them in the results section.

Results

In this section we present several statistical analyses aimed at addressing the following research questions regarding the planning of verbal responses in spontaneous conversation (corresponding results sections below between brackets):

(a) Are the characteristics of pre-utterance inbreaths related to the scope of language planning? (Sections Pre-utterance Inbreaths and Answer Duration, and Inbreath Characteristics and Answer Duration).
(b) Are responses preceded by an inbreath delayed with respect to end of the interlocutor's turn compared to responses spoken on residual breath? (Section Pre-utterance Inbreaths and Answer Latency).

(c) What is the most common timing of pre-utterance inbreaths relative to the end of the interlocutor's turn? (Section Timing of Answerer's Inbreaths Relative to Question Ends).

Pre-Utterance Inbreaths and Answer Duration

As mentioned above, only 53.2% of the answers to questions were preceded by an inbreath. We first examined whether the presence or absence of pre-utterance inbreaths is related to the duration of the answer. Because pre-utterance inbreaths could also be affected by the preference status of the answer, and by the presence of disfluencies in the answer, we first fit two reduced logistic mixed-effects regression models with either of these two variables as fixed predictors, speaker as a random factor, and the presence of a pre-utterance inbreath as the response. The preference status of the answer did not significantly improve the fit of a null model ($p = 0.52$), and was therefore dropped from subsequent analyses. On the other hand, the presence of disfluencies in the answer provided a highly statistically significant improvement over the null model [$\chi^2_{(1)} = 17.21, p < 0.0001$], indicating that pre-utterance inbreaths are more likely before answers containing one or more hesitations. Interestingly, a model including the presence of disfluencies in the answer plus answer duration compared favorably to a model including the presence of disfluencies only [$\chi^2_{(1)} = 6.38, p < 0.05$], and indicated that pre-utterance inbreaths are more likely the longer the answer [$\beta = 0.35, z = 2.31, p < 0.05$]. This is illustrated in **Figure 3**, which shows the percentage of pre-utterance inbreaths as a function of answer duration. It should be noted that in the full model the β coefficient for the presence of disfluencies in the answer was not statistically significant from 0 ($\beta = 0.76, z = 1.78, p = 0.07$), perhaps due to the fact that this variable and answer duration, the other fixed predictor, were moderately correlated ($r = 0.59$). In the same way, adding the presence of disfluencies in the answer to a model with answer duration as the only fixed predictor did not result into a statistical improvement [$\chi^2_{(1)} = 3.24, p = 0.07$]. Thus, the relationship between answer duration and pre-utterance inbreaths cannot be explained away by the correlation between answer duration and the presence of disfluencies in the answer. Instead, it appears that answer duration is a better predictor of whether a pre-utterance inbreath is present than the fluency of the answer.

Inbreath Characteristics and Answer Duration

We then examined if, within the group of answers preceded by an inbreath, answer duration was related to inbreath depth (in speaker-normalized z-scores) and inbreath duration (in seconds; mean = 0.887 s, median = 0.72 s). We first fitted reduced models with inbreath depth and inbreath duration as responses, and

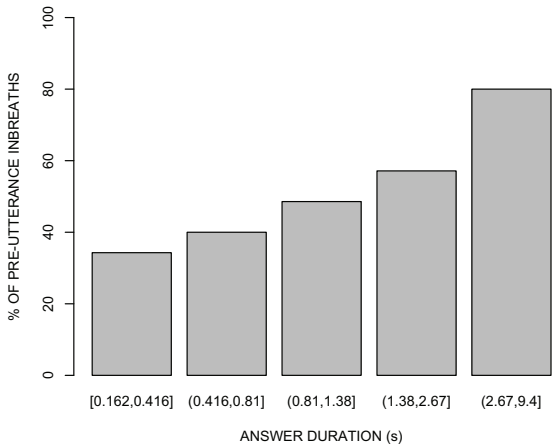


FIGURE 3 | Percentage of pre-utterance inbreaths as a function of answer duration (s) for five quantiles of approximately equal size (*n* = 35 for the lowest quantile, and *n* = 34 for all other quantiles).

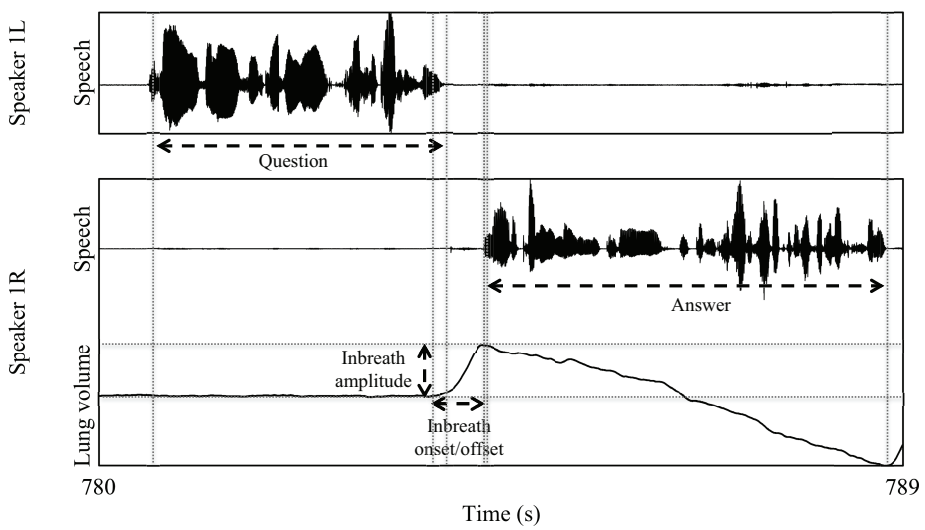


FIGURE 2 | Illustration of our measurements in a question and answer sequence exhibiting an inbreath before the answer.

either the preference status of the answer or the presence of disfluencies in the answer as fixed predictors, and observed that none of the fixed predictors was statistically related to any of the two responses ($p > 0.05$ in all comparisons with a null model). Adding answer duration to the null models did not improve its fit either for neither of the two response variables [inbreath duration: $\chi^2_{(1)} = 0.16$, $p = 0.69$; inbreath depth: $\chi^2_{(1)} = 0.2$, $p = 0.64$]. Moreover, visual inspection of the data indicated that this lack of statistical relationships was not due to outliers. Thus, contrary to previous findings (Winkworth et al., 1995; Whalen and Kinsella-Shaw, 1997; Fuchs et al., 2013), we did not observe any statistical relationship between utterance duration and the amplitude and duration of pre-utterance inbreaths.

Pre-Utterance Inbreaths and Answer Latency

The main question that we wanted to answer in this study concerns whether answerers produce inbreaths in anticipation of question ends in order to produce answers without substantial delays (compared to answers not preceded by a pre-utterance inbreath), or if pre-utterance inbreaths occur close to turn ends, rendering responses later than those without preceding inbreaths. In order to investigate this, we first fitted reduced regression models with answer latency as the response, and either the preference status of the answer or the presence of disfluencies in the answer as a fixed predictor. None of these factors improved the null model ($p > 0.05$ in both cases). Because longer answers may take longer to plan, we also fitted a reduced model with answer duration as the fixed predictor. In this case, there was a statistical improvement over the null model [$\chi^2_{(1)} = 4.61$, $p < 0.05$]. Interestingly, adding the occurrence of a pre-utterance inbreath greatly improved the fit of the model [$\chi^2_{(1)} = 11.2$, $p < 0.001$]. As illustrated in **Figure 4**, answers preceded by an inbreath were substantially more delayed with respect to the end of the question than answers not preceded by an inbreath. The mean, standard deviation, median, and estimated mode for answers preceded and not preceded by an inbreath are shown in **Table 1** (the mode of answer latency and other continuous variables was estimated with the function density() in R set to default parameters).

Timing of Answerer's Inbreaths Relative to Question Ends

Figure 5 shows a histogram of the timing of answerer inbreaths relative to question ends. The mean and median of this measure were respectively −309 and −56 ms. Fitting the data with a continuous density function in R, the mode of the distribution was estimated at 15 ms, that is, briefly after the end of the question. The example in **Figure 2**, in which the answerer's inbreath is aligned close to the end of the question, is therefore representative of the most frequent cases in our data. However, there were also cases with much earlier timings, sometimes with inbreaths starting a second or more in advance of the question end. Individual inspection of such cases suggested that some of them may not have been primarily designed for speech. For instance, some of these early inbreaths were produced immediately after the end of a long turn, and were therefore likely to be conditioned more by the previous than the upcoming utterance (i.e., the answer to the question).

TABLE 1 | Mean and standard deviation (SD), median, and estimated modal answer latencies relative to question ends for answers preceded and not preceded by an inbreath.

	Answer latency		
	Mean (*SD*)	**Median**	**Estimated mode**
No inbreath	459 (659) ms	347 ms	100 ms
Inbreath	998 (1008) ms	823 ms	576 ms

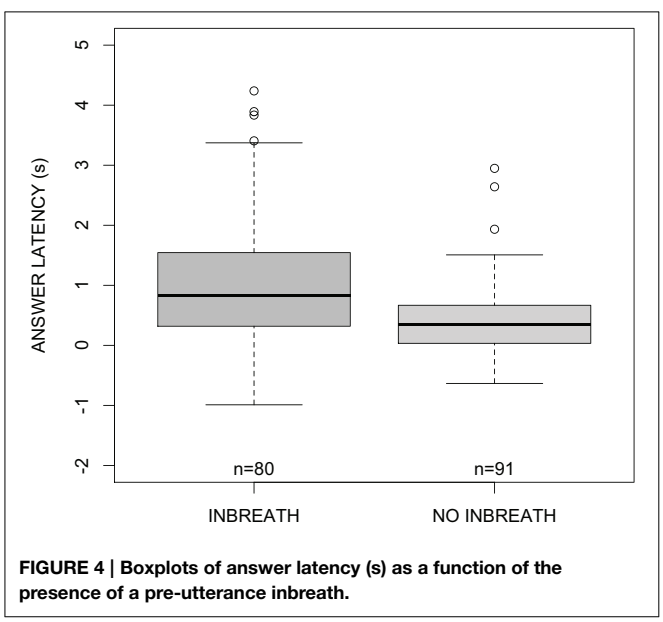


FIGURE 4 | Boxplots of answer latency (s) as a function of the presence of a pre-utterance inbreath.

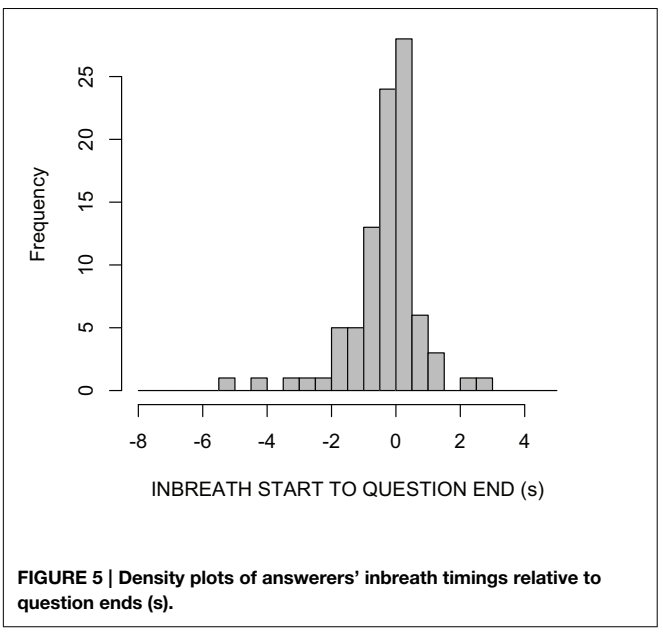


FIGURE 5 | Density plots of answerers' inbreath timings relative to question ends (s).

It is possible that inbreaths that are intended as semiotic signals, for instance announcing an upcoming dispreferred or disfluent answer, may tend to be produced in the clear rather than in overlap with the interlocutor's turn end. To investigate this, we fitted regression models with the distance from inbreath start to question end as the response, and either the preference status of the answer or the presence of disfluencies in the answer as fixed predictors. None of these two predictors provided an improvement over the null model ($p > 0.5$ in both cases). It therefore appears that the clustering of inbreath starts close to question ends is not related to the preference status or the fluency of the response.

In our data, therefore, the most typical timing of inbreaths, as captured by median and modal values, is strikingly close to the question end. This finding suggests that answerers tend to coordinate the onset of their vocal behavior, in this case an inbreath, with the end of their interlocutors' turn. However, we need to rule out an alternative interpretation, namely that the frequent alignment of inbreaths with question ends was simply caused by our annotation criteria. Recall that we annotated answerer's inbreaths only if they occurred between the beginning of the question and the beginning of the answer, that is, if they occurred either in overlap with the question or during the question-answer transition (see **Figure 2** above). In a scenario in which the timing of inbreaths is random and the duration of the considered time interval is constant, we would expect a uniform distribution of inbreath timings throughout the considered time interval. However, because the considered time interval in our data was variable, it was not possible to determine the expected distribution of inbreath timings under the random timing hypothesis in a straightforward way. In order to estimate such distribution, we generated 1000 distributions of random inbreath timings within the considered time intervals in our data, and compared them with the observed distribution of inbreath timings. Because the minimum inbreath duration in our data was 210 ms, we allowed the random inbreath timings to occur randomly anywhere between the beginning of each question in the data, and 210 ms before the beginning of its answer.

Figure 6 shows 1000 overlaid density plots representing the randomly generated distributions (thin solid lines), along with the observed distribution (dashed line). On visual inspection, the distributions of random timings appear to have lower measures of central tendency than the observed distribution. In fact, all of the medians of the random-timing distributions were lower than the observed mode; all of the modes of the random-timing distributions were lower than the observed mode; and only 35 out of the 1000 means of random-timing distributions were equal or higher than the observed mean. Based on these proportions, the estimated probabilities that the observed median, mode, and mean were generated by a distribution of random timings are very low (i.e., median: $p < 0.001$; mode: $p < 0.001$; mean: $p < 0.035$). This suggests that the frequent alignment between answerer inbreaths and question ends observed in our data is unlikely to be random, and that it is likely to be a genuine index of coordination between questioners and answerers.

Discussion

Our findings can be summarized as follows. First, answerers' breathing behavior in question-answer sequences in conversation is related to answer length, and this relationship cannot be explained by either the preference status of the answer or the presence of disfluencies occurring in the answer. Long answers have a greater probability of being preceded by an inbreath than short answers. In contrast, we did not find any relation between answer length and inbreath characteristics

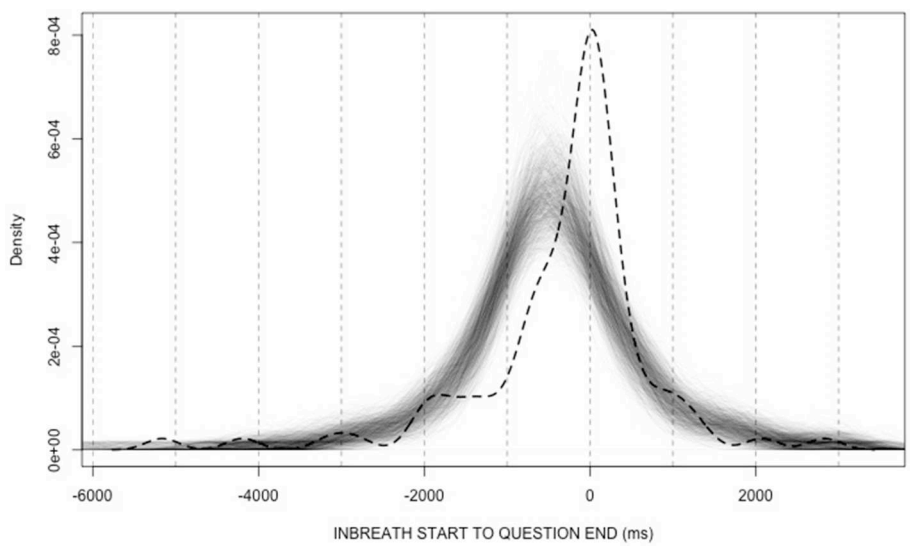


FIGURE 6 | Density plots of answerers' inbreath timings relative to question ends. The dashed curve represents the distribution in our data (see **Figure 5**), while the overlaid thin lines represent randomly generated distributions.

such as duration and amplitude, as reported in previous studies (Winkworth et al., 1995; Whalen and Kinsella-Shaw, 1997; Fuchs et al., 2013; Rochet-Capellan and Fuchs, 2013). Note, however, that Winkworth et al. (1995), who, like us, studied spontaneous conversations, pooled turn-internal inbreaths and inbreaths at turn transitions together, whereas we focused on question-answer sequences always involving a predictable floor transfer.

Second, answer latencies are significantly longer when an inbreath precedes the answer. The most frequent timing for answers preceded by an inbreath was around 600 ms after the question end, while the most frequent timing for answers not preceded by an inbreath was 100 ms. Third, we found that, despite significant variability in the data, there was a clear tendency for answerers to inhale briefly after the end of their interlocutors' questions, with an estimated modal offset of 15 ms. We have also shown that this tendency is not merely a chance outcome due to the durational properties of the question and answer sequences in our data. Interestingly, this finding appears to be congruous with earlier findings by McFarland (2001) and Rochet-Capellan and Fuchs (2014). McFarland (2001) studied breathing kinematics in a number of conditions, including unscripted dialogue. Using a cross-correlation method, he observed that turn exchanges were associated with a high number of significant correlations between the breathing signals of the conversation participants. These correlations were sometimes negative, indicating an anti-phase coupling, and sometimes positive, indicating an in-phase relationship. Rochet-Capellan and Fuchs (2014), also using spontaneous conversation data, did not observe a general interpersonal coordination of breathing behavior over whole conversations, but did observe specific coordinative patterns in shorter time-windows when participants exchanged turns. Although we have not examined the breathing behavior of questioners in our data, it is reasonable to assume that they often took an inbreath soon after finishing their turns. Since answerers in our data tended to inhale close to the end of questions, it is quite plausible that the breathing cycles of questioners and answerers often were in an in-phase relationship within the temporal region of the turn transition.

Our analysis of preference revealed that this factor was not a major source of variability in the breathing behavior of responders in our data. This stands in contrast with the findings of Kendrick and Torreira (2015), who found that dispreferred responses in a corpus of telephone calls tend to be preceded by an inbreath more often than preferred responses. This is perhaps due to the fact that the present study considered all pre-utterance inbreaths registered through inductive plethysmography, whereas Kendrick and Torreira could only have access to those that were audible in their acoustic data. Another possible explanation is that Kendrick and Torreira focused on a restricted number of conversational actions (responses to invitations, offers, and requests) that could take on variable linguistic forms, whereas we focused on specific linguistic forms (polar and wh-questions as defined in section Coding and Measures) that accomplished an unspecified number of actions. Finally, it is also possible that inbreaths often act as preference markers in telephone conversations, but not in face-to-face interactions. In telephone conversations, interactants do not see each other, and can only use acoustic information in order to communicate. Moreover, since speakers in telephone calls typically hold their telephones close to their lips and ears, subtle mouth noises such as inbreaths and clicks may be more efficient communicative signals in telephone calls than in face-to-face conversation.

We turn now to the interpretation of our findings. The main goal of this study was to evaluate two competing hypotheses concerning the most typical time course of language planning and production during conversational turn-taking. A model in which the articulation of one's turn relies on early prediction of turn-end timing and disregards turn-final cues (cf. De Ruiter et al., 2006) posits that listeners typically estimate the end of the incoming turn well in advance of the turn end (i.e., over 500 ms; Magyari and de Ruiter, 2012, and that they plan and launch their response in anticipation of that predicted time point. If we take into account that pre-utterance inbreaths usually last several hundred milliseconds (over 800 ms on average in our data), this model predicts that listeners will produce them in overlap with the incoming turn, so as to be able to start speaking close to the estimated turn end. On the other hand, a model consisting of early planning of content and late triggering of articulation based on turn-final cues, as discussed in Heldner and Edlund (2010), predicts that listeners will produce pre-utterance inbreaths close to the end of the interlocutor's turn, and that answers preceded by an inbreath will be delayed compared to answers produced on residual breath. Our data collected via inductive plethymosgraphy indicate that the most typical moment in which responders take a pre-utterance inbreath is briefly after the end of the question, not several hundred milliseconds in advance of its end. As a consequence of this, answers preceded by an inbreath were delayed relative to answers which were not. Our findings thus favor a model based on early prediction of content plus late triggering of articulation based on information present close to turn ends. Although we cannot discard the possibility that interlocutors use projection of turn-end timing in specific situations, our observational data suggest that late launching of vocal behavior is a more common strategy.

Since activation of the internal intercostal muscles, which are usually involved in breathing activity, requires minimally 140 ms (Draper et al., 1960), and inbreaths typically occur a few ms after the question end, we can infer that inbreath preparation for answers most often starts during the last syllable, word, or foot of the question, where phrase-final prosodic cues (e.g., final lengthening, final pitch accents, and boundary tones in a language like Dutch) and possibly other phonetic cues to turn ends (Local and Walker, 2012) become manifest. Interestingly, answers not preceded by an inbreath most frequently occurred 100 ms after the end of the question. Allowing for a minimal vocal response time of 210 ms (Fry, 1975), it can be surmised that the articulation of such answers is launched roughly at the same time as pre-utterance inbreaths when these are present. Our data therefore suggest that the launching of physical responses at turn transitions, either in the form of pre-utterance inbreaths or speech proper, typically occurs in reaction to information present in the last portion of the interlocutor's utterance. **Figure 7** shows two

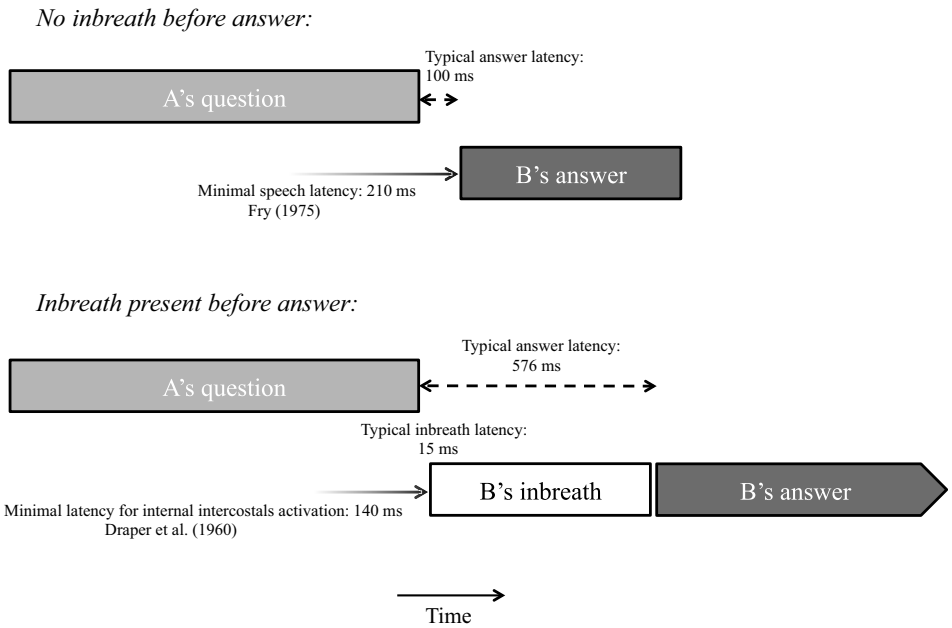


FIGURE 7 | Schematic illustration of two typical time courses of vocal behavior in question-answer sequences, along with minimal response latencies reported in previous literature.

typical time courses for vocal responses to a question in schematic form.

The fact that answerers tend to inhale more often before long answers, together with the typical alignment of inbreaths with question ends, implies that some amount of conceptual planning involving the size of the answer must already take place in overlap with the incoming question. This finding is consistent with recent EEG evidence that response preparation starts well in overlap with the incoming turn (Bögels et al., 2014). The claim that production planning significantly overlaps with comprehension processes is nevertheless puzzling, because it implies dual tasking using much of the same neural circuitry (e.g., Indefrey and Levelt, 2004; Menenti et al., 2011; Hagoort and Indefrey, 2014) and an intensive sharing of attentional resources (cf. Jongman et al., 2015). One can only speculate about how this may be possible, for example, by a rapid switching of resources between the two processes, with a gradual increase of allotted time-share to production.

The considerations on the time course of language production in conversational turn-taking presented above are based on the most typical values observed in our data, and on minimal response latencies reported in previous research. Importantly, however, we also observed a significant amount of variability in breathing and answer latencies, with relatively long overlaps and gaps accounting for a substantial portion of the data. Under the two-stage production mechanism outlined above (i.e., early planning of content overlapping with the interlocutor's turn, plus late launching of articulation based on incoming turn-final cues), such non-smooth turn transitions require further explanation. Such cases could arise when either early language planning or the launching of articulation based on turn-final cues are not carried out optimally. For instance, one common cause of speech overlap routinely mentioned in the Conversational Analysis literature (e.g., Jefferson, 1986) is that turns may contain several potential ends (i.e., transition relevance points, or TRPs) within them (e.g., "Are you coming later? To the party?"), and that listeners may time their turn with respect to one of the non-final possible turn ends (e.g., the word "later" in the previous example). Launching articulation without waiting to hear a silence at the end of the interlocutor's turn is, in fact, what our data suggest, and what our model predicts.

In cases of long inbreath latencies, the responder may not have been able to plan the initial stages of her turn (e.g., conceptual planning) early enough to determine whether she needs to take an inbreath before her turn, and launch it in response to the interlocutor's turn-final cues. This may be due to a low attentional level on the part of the speaker, or to the interlocutor's turn being unclear until its very end. In cases in which the speaker is able to complete the initial stages of language production in time to provide a smooth response, but not the later stages (e.g., phonological encoding of the beginning of her turn), she could still take an early inbreath upon identification of the turn-final cues in the interlocutor's turn, and then use her inbreath, which may stretch for several hundred milliseconds, as a buffer through which to complete the planning of the utterance.

We hope to have shown that the study of breathing can shed new and interesting light on the underlying mechanisms involved in turn-taking. The current study is limited to question-answer contexts in which answers are always produced in response to a

question. We think that our conclusions regarding the answerer's breathing behavior can be expected to be valid in turn-taking contexts involving readiness to respond on the part of one of the interlocutors. However, further research should explore other conversational contexts in which floor changes may be optional (i.e., end of conversational sequences), subject to increased competition for the floor (e.g., multi-party conversation) or involving highly predictable first turns (cf. Magyari and de Ruiter, 2012), since different production mechanisms might be used in different situations. It would also be interesting to relate the breathing signal to other early signals of speech preparation obtained by direct measurement of the vocal organs via ultrasound (Drake et al., 2014; Palo et al., 2014; Schaeffler et al., 2014) or other instrumental techniques such as electromagnetic articulography. We believe this is a rich field that should be further explored.

Acknowledgments

This work was supported by an ERC Advanced Grant (269484 INTERACT) to SL. We would like to thank Ronald Fischer, Mart Lubbers, and Imke van Brink for their assistance during the collection and processing of the data. We are grateful to Marisa Casillas and Sean Roberts for helpful comments on earlier versions of this article. We are also grateful to the members of the IFL and Dialogue projects at the Max Planck Institute for Psycholinguistics for extensive discussion of this work.

References

Bailly, G., Rochet-Capellan, A., and Vilain, C. (2013). "Adaptation of respiratory patterns in collaborative reading," in *Proceedings of Interspeech* (Lyon), 1653–1657.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear mixed-Effects Models Using Eigen and S4*. R Package Version 1. Available online at: http://CRAN.R-project.org/package=lme4.Bögels.

Bögels, S., Magyari, L., and Levinson, S. C. (2014). "Neural correlates of speech preparation in interactive turn-taking: An early start?" in *Society for the Neurobiology of Language Meeting 2014* (Amsterdam).

Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *J. Phon.* 31, 251–276. doi: 10.1016/S0095-4470(03)00007-X

De Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

Drake, E., Schaeffler, S., and Corley, M. (2014). "Articulatory effects of prediction during comprehension: an ultrasound tongue imaging approach," in *Proceedings of the 10th International Seminar in Speech Production* (Cologne), 98–101.

Draper, M. H., Ladefoged, P., and Whitteridge, D. (1960). Expiratory pressures and airflow during speech. *Br. Med. J.* 1, 1837–1842. doi: 10.1136/bmj.1.5189.1837

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *J. Pers. Soc. Psychol.* 23, 283–292. doi: 10.1037/h0033031

Fry, D. B. (1975). Simple reaction-times to speech and non-speech stimuli. *Cortex* 11, 355–360. doi: 10.1016/S0010-9452(75)80027-X

Fuchs, S., Petrone, C., Krivokapic, J., and Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *J. Phon.* 41, 29–47. doi: 10.1016/j.wocn.2012.08.007

Griffin, Z. M., and Bock, K. (2000). What the eyes say about speaking. *Psychol. Sci.* 11, 274–279. doi: 10.1111/1467-9280.00255

Hagoort, P., and Indefrey, P. (2014). The neurobiology of language beyond single words. *Ann. Rev. Neurosci.* 37, 347–362. doi: 10.1146/annurev-neuro-071013-013847

Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversation. *J. Phon.* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002

Indefrey, P., and Levelt, W. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001

Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Hum. Stud.* 9, 153–183. doi: 10.1007/BF00148125

Jongman, S. R., Roelofs, A., and Meyer, A. S. (2015). Sustained attention in language production: an individual differences investigation. *Q. J. Exp. Psychol.* doi: 10.1080/17470218.2014.964736. [Epub ahead of print].

Kendrick, K. H., and Torreira, F. (2015). The timing and construction of preference: a quantitative study. *Discourse Process.* doi: 10.1080/0163853X.2014.955997. [Epub ahead of print].

Levelt, W., Roelofs, A., and Meyer, A. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–37. doi: 10.1017/S0140525X99001776

Local, J., and Walker, G. (2012). How phonetic features project more talk. *J. Int. Phon. Assoc.* 42, 255–280. doi: 10.1017/S0025100312000187

Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *J. Speech Lang. Hear. Res.* 44, 128–143. doi: 10.1044/1092-4388(2001/012)

Menenti, L., Gierhan, S., Segaert, K., and Hagoort, P. (2011). Shared language overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychol. Sci.* 22, 1173–1182. doi: 10.1177/0956797611418347

Palo, P., Schaeffler, S., and Scobbie, J. M. (2014). "Pre-speech tongue movements recorded with ultrasound," in *Proceedings of the 10th International Seminar in Speech Production* (Cologne), 300–303.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org/.

Rochet-Capellan, A., and Fuchs, S. (2013). "The interplay of linguistic structure and breathing in German spontaneous speech," in *Proceedings of Interspeech 2013*, 1128–1132.

Rochet-Capellan, A., and Fuchs, S. (2014). Take a breath and take the turn: how breathing meets turns in spontaneous dialogue. *Philos. Trans. R. Soc. B* 369:20130399. doi: 10.1098/rstb.2013.0399

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.2307/412243

Schaeffler, S., Scobbie, J. M., and Schaeffler, F. (2014). "Measuring reaction times: vocalisation vs. articulation," in *Proceedings of the 10th International Seminar in Speech Production* (Cologne), 379–382.

Schegloff, E. A. (2006). *Sequence Organization in Interaction*, Vol. 1. Cambridge, UK: Cambridge University Press.

Stivers, T., and Enfield, N. J. (2010). A coding scheme for question–response sequences in conversation. *J. Pragmat.* 42, 2620—2626. doi: 10.1016/j.pragma.2010.04.002

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Wells, B., and MacFarlane, S. (1998). Prosody as an interactional resource: turn-projection and overlap. *Lang. Speech* 41, 265–294.

Whalen, D. H., and Kinsella-Shaw, J. M. (1997). Exploring the relationship of inspiration duration to utterance duration. *Phonetica* 54, 138–152. doi: 10.1159/000262218

Winkworth, A., Davis, P., Adams, R., and Ellis, E. (1995). Breathing patterns during spontaneous speech. *J. Speech Lang. Hear. Res.* 38, 124. doi: 10.1044/jshr.3801.124

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN: a professional framework for multimodality research," in *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation* (Genoa).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 March 2015
doi: 10.3389/fpsyg.2015.00250

The intersection of turn-taking and repair: the timing of other-initiations of repair in conversation

Kobin H. Kendrick *

Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

The transitions between turns at talk in conversation tend to occur quickly, with only a slight gap of ∼100–300 ms between them. This estimate of central tendency, however, hides a wealth of complex variation, as a number of factors, such as the type of turns involved, have been shown to influence the timing of turn transitions. This article considers one specific type of turn that does not conform to the statistical trend, namely turns that deal with troubles of speaking, hearing, and understanding, known as other-initiations of repair (OIR). The results of a quantitative analysis of 169 OIRs in face-to-face conversation reveal that the most frequent cases occur after gaps of ∼700 ms. Furthermore, OIRs that locate a source of trouble in a prior turn specifically tend to occur after shorter gaps than those that do not, and those that correct errors in a prior turn, while rare, tend to occur without delay. An analysis of the transitions before OIRs, using methods of conversation analysis, suggests that speakers use the extra time (i) to search for a late recognition of the problematic turn, (ii) to provide an opportunity for the speaker of the problematic turn to resolve the trouble independently, and (iii) to produce visual signals, such as facial gestures. In light of these results, it is argued that OIRs take priority over other turns at talk in conversation and therefore are not subject to the same rules and constraints that motivate fast turn transitions in general.

Keywords: conversation analysis, turn-taking, timing, delay, other-initiated repair, self-repair, preference

OPEN ACCESS

Edited by:
Marcela Pena, Catholic University of Chile, Chile

Reviewed by:
Pilar Prieto, Institució Catalana de Recerca i Estudis Avançats – Universitat Pompeu Fabra, Spain
Asli Ozyurek, Radboud University Nijmegen, Netherlands

***Correspondence:**
Kobin H. Kendrick, Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525 XD, Netherlands
kobin.kendrick@mpi.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *16 October 2014*
Accepted: *18 February 2015*
Published: *12 March 2015*

Citation:
Kendrick KH (2015) The intersection of turn-taking and repair: the timing of other-initiations of repair in conversation. Front. Psychol. 6:250. doi: 10.3389/fpsyg.2015.00250

Introduction

In conversation opportunities to participate are organized by a system of turn-taking (Sacks et al., 1974). The rules and constraints of the turn-taking system conspire to minimize the duration of transitions between turns. But not all transitions are in fact minimal. The transitions before turns that deal with troubles of speaking, hearing, and understanding, known as other-initiations of repair (OIRs; e.g., "what?," "who?," "what'd you mean?"), have been reported by Schegloff et al. (1977) to be longer than those before other turns. How much longer, however, remains an open question. The first goal of this investigation is therefore to verify and refine this observation though a quantitative analysis of the timing of OIR, using responses to polar question as a point of comparison. Schegloff et al. (1977) argue that next speakers withhold OIRs to provide an opportunity for current speakers to resolve the trouble via self-repair. Whether this exhausts the possible explanations for delay before OIR is unclear, however. The second goal of the investigation is therefore to look inside the transition spaces before OIRs, using conversation-analytic methods to discover and describe what speakers use them to accomplish. As background to this, I begin with a discussion of the timing of turn-taking in general and the timing of OIR more specifically.

The Timing of Turn-Taking

Previous research on the timing of turn-taking has shown that the transitions between turns in conversation most frequently occur with only minimal gaps and overlaps. First documented systematically in a series of meticulous conversation-analytic studies (Jefferson, 1973, 1983a,b, 1984, 1986; Sacks et al., 1974), the timing of transitions between turns has subsequently been investigated primarily through quantitative methods. In general, quantitative studies have taken one of two approaches, either examining all transitions within a corpus, irrespective of turn type (e.g., Wilson and Zimmerman, 1986; Heldner and Edlund, 2010), or analyzing just one type of transition, that between questions and answers (e.g., Stivers et al., 2009; Kendrick and Torreira, 2014). A comparison between the timing of transitions between questions and answer and that of a random sample of transitions in a corpus of Dutch conversation revealed no statistically significant difference (Stivers et al., 2009). This suggests that the timing of question–answer sequences can be used as a proxy for a typical turn transition in conversation. The results of these studies generally converge, indicating that the most frequent transitions between turns occur with a slight gap (cf. Jefferson, 1984, p. 18), on the order of 100–300 ms.

This estimate of central tendency has clear implications for psycholinguistic models of turn-taking. As Levinson (2013) points out, psycholinguistic research has shown that speakers require a minimum of 600 ms to plan even a single word (e.g., in a picture-naming task; Indefrey and Levelt, 2004; Indefrey, 2011). Thus the average gap between turns does not provide enough time for a speaker to prepare even a simple next turn. Therefore, Levinson argues, next speakers must anticipate the ends of turns, and begin to plan the next turn well before the current one is complete, in full agreement with arguments by Sacks et al. (1974) and Schegloff (1987).

But an estimate of central tendency is by definition a simplification, a single value that ideally represents a more complex distribution. Research has also examined sources of complex and systematic variation in the timing of turn-taking, especially in question–answer sequences. The language and culture of the speakers, the deployment of gaze, and the type and modality of the response all have been shown to influence the timing of responses to polar questions (Stivers et al., 2009). Question type is also relevant; responses to polar questions are generally faster than responses to content questions (Strömbergsson et al., 2013). Studies that explore variation in the timing of turn transitions are important because they reminds us of the diversity of turn types and hence turn transitions in conversation.

The Timing of Other-Initiations

At least one type of turn that does not conform to global generalizations about the timing of turn-taking has been identified in the conversation-analytic literature. Turns that deal with troubles of hearing or understanding prior turns (e.g., "what?," "who?," "what'd you mean?"), referred to as OIRs, have been reported to be systematically withheld (Schegloff et al., 1977). The transition between the turn that contains the trouble (e.g., an error or a word the next speaker does not understand), referred to as the trouble-source turn, and the other-initiation of repair has been observed to be longer than other turn transitions (Schegloff et al., 1977; Robinson, 2006).

Based on a systematic qualitative analysis, Schegloff et al. (1977, p. 374) observed that OIRs "regularly are withheld" by speakers and therefore "occur after a slight gap." Although the report includes numerous cases of OIRs that occur after a slight gap (and many that do not), it leaves basic questions unanswered, such as the frequency with which speakers withhold OIRs and the durations of the gaps that precede them.

A second report of the phenomenon, made in passing by Robinson (2006, p. 153), contains more detailed information. Based on 32 cases of OIR in telephone conversations, Robinson observed that the median delay was between 0.1 and 0.2 s. This finding is not conclusive, however, for two reasons. The first concerns the method of timing. Rather than measure the duration of gaps and pauses objectively (e.g., using a computer), conversation analysts typically employ a relative method of timing, one that reflects the analyst's perception of time (Hepburn and Bolden, 2013), a method that has been shown to overestimate objectively measured time systematically (Roberts and Robinson, 2004; Kendrick and Torreira, 2014). The second issue is the lack of an explicit comparison between the timing of OIRs and the timing of other turns in the same conversations. To conclude that OIRs are delayed systematically, one must establish not only that gaps before them are long, but more importantly that they are *longer* than gaps before other turns. Thus while Robinson's finding supports the claim by Schegloff et al. (1977), the frequency with which speakers do or do not withhold OIRs and the precise timing of the gaps that precede them remain open questions.

The Practices of Other-Initiation

The observation of systematic variation in the timing of turn types (e.g., responses to questions, noted above) points to further questions. Is variation in the timing of OIRs also systematic? Do different types of OIRs, like different types of responses to questions, tend to occur after relatively shorter or longer transitions?

The inventory of OIR practices in English is relatively well described (see Schegloff et al., 1977; Benjamin, 2013; Kitzinger, 2013; Kendrick, in press). A basic distinction is made between OIRs that pinpoint a specific source of trouble in a prior turn and those that do not. Open OIRs such as "what?" or "sorry?" indicate that the speaker has encountered a trouble but do not specify a particular source (Drew, 1997). In contrast, OIRs that repeat all or part of a prior turn (e.g., "they're what?") or request category-specific information (e.g., "who?"), among other possibilities, specifically locate the source of the trouble. According to Schegloff et al. (1977), there is a preference for more specific ('stronger') over less specific ('weaker') OIRs, such that speakers should, for example, use a specific OIR over an open OIR if possible (cf. Svennevig, 2008). This raises the question of whether the timing of open and specific OIRs differs systematically and whether it provides evidence for or against the preference for specificity.

A further distinction is whether an OIR constitutes a correction or not (Schegloff et al., 1977; Jefferson, 1987). An OIR can proffer a replacement of a trouble source as a candidate

solution to a trouble of hearing or understanding (e.g., A: "she got so mad." B: "Pam's mother?" A: "Mm hm."). Or it can assert a replacement as correction of an error (e.g., A: "she's eating the Butterworth diet." B: "Butterfield." A: "Butterfield."). Correction by someone other than the speaker of the error, known as other-correction, has been argued to be a dispreferred alternative to self-correction. Schegloff et al. (1977) observed that other-corrections tended to exhibit special marking and special positioning (e.g., the qualification of epistemic stance or delay within a turn or sequence) that revealed them to be dispreferred actions. But counter-examples to this generalization are not uncommon (see, e.g., Jefferson, 1987, p. 87), and a recent survey of other-correction in Finnish failed to identify special markings of the type that Schegloff et al. described (Haakana and Kurhila, 2009). What, then, of the timing of other-corrections? Are the transitions before other-corrections longer than other turns, as their putative status as dispreferred actions predicts?

The Motivations for Delay Before Other-Initiations

A final set of questions concerns the motivations for and consequences of delay before OIR. Conversation analysts have argued that the timing of OIRs has a socio-interactional basis. A possible next speaker who encounters a problematic turn withholds an OIR in order to provide the current speaker with an opportunity to resolve the trouble on his or her own (Schegloff et al., 1977). If a next speaker does not understand a question, for example, he or she might not immediately reply with "what'd you mean?" but might first wait for a moment to give the current speaker an opportunity to repair the question independently, using practices of self-initiated repair. In this way, the self-initiation of repair takes precedence over the other-initiation of repair, one aspect of a principle known as the preference for self-correction (Schegloff et al., 1977). The timing of the transition between a trouble-source turn and an other-initiation of repair is thus seen as a locus for the management of basic social relations, between self and other.

But a priori one might propose complementary or alternative explanations for the phenomenon of delay before OIR. The private processes that speakers necessarily engage in to hear and understand a turn at talk occur in real time and, just like other actions, take time to complete. The high frequency of transitions with minimal gaps suggests that these processes very often occur quickly. But might they not take longer under certain conditions, say, when a next speaker has failed to hear, or understand a prior turn? Is the delay before an OIR for the benefit of the speaker of the trouble, to provide an opportunity for self-initiated repair, or is it (also) for the benefit of the recipient of the trouble, to allow a search for and a possible recovery of a hearing and understanding of the turn that may permit the sequence to move forward without repair?

One might also look to the embodied actions of the participants for an explanation. The private processes that participants engage in surely take time to complete, but so too do the physical actions they perform. A withdrawal of gaze from the speaker of a trouble-source turn, a rotation of the head and body to face a trouble-source speaker, a meaningful deployment of facial muscles into a gesture of puzzlement—embodied actions such as these all take time to produce. Through the filter of a text transcript, a silence between turns at talk, whether long, or short, can look like an absence of action. But the long tradition of multimodal conversation analysis (Goodwin, 1980, 1981; Schegloff, 1998; Lerner, 2003; Mondada, 2006, 2007; Rossano, 2012; inter alia) has shown that action does not necessarily end with a turn at talk. What, then, do speakers *do* in the transition spaces before OIR?

The Current Investigation

The current investigation combines conversation-analytic and quantitative methods to address the following questions about the timing of OIR.

(1) Are the transition spaces before OIRs systematically longer than those before other turn types, such as answers to questions?
(2) Are the transition spaces before some types of OIRs systematically longer than those before other types?
(3) What are the motivations for the expansion of transition spaces before OIRs? That is, what do speakers use the extra time to accomplish?

The investigation uses conversation analysis to identify and analyze OIRs and question–answer sequences, as well as to discover and describe a number of uses that speakers have for the expanded transition spaces before OIRs. The measurements of transition spaces and the comparisons of the distributions are done quantitatively. At the end of the article I return to the questions of the motivations for and consequences of delay before OIRs and consider the relationship between repair and the turn-taking system.

Materials and Methods

Data

The data for the investigation came from video-recordings of naturally occurring English conversation between friends and family members engaged in a variety of activities (e.g., chatting, playing games, preparing food, eating dinner). The corpus consisted of 19 recordings, with a total duration of 9 h and 20 min, and included native speakers of English from the U.S., Canada, and the U.K. Informed consent was obtained from all participants.

Identification of OIR

All cases of OIR were systematically identified in the corpus, using the methods of conversation analysis and drawing on previous research on OIR (Schegloff et al., 1977; Jefferson, 1987; Schegloff, 1992; Robinson, 2006; Egbert et al., 2009; Robinson and Kevoe-Feldman, 2010; Benjamin and Walker, 2013). OIRs were distinguished from formally similar practices that do not initiate repair as an action (see Schegloff, 1997, for examples). It is well known that the practices of OIR can be used to display surprise or ritualized disbelief (Selting, 1988; Wilkinson and Kitzinger, 2006). In the case of repeats of a

trouble source, a clear boundary between such cases and those that additionally or alternatively display surprise or disbelief has not yet been identified in the literature. Such cases were therefore included in the investigation. The types of OIRs identified in the corpus are given in **Table 1** along with examples (for a more detailed report of the distribution of OIR in English conversation, see Kendrick, in press). A total of 222 cases of OIR were identified, for an average rate of one every 2.5 min.

Open and Specific OIRs

Other-initiations of repair differ in the specificity with which they locate the source of trouble in the prior talk (Schegloff et al., 1977). Open OIRs indicate that the speaker has encountered trouble with the prior talk but they do not specify a particular source (e.g., "what?"; see Drew, 1997; Benjamin, 2013). In contrast, specific OIRs locate a particular component in the prior talk as the trouble source (e.g., "who?" or "she did what?"). The practices listed **Table 1**, other than those designated as open, were analyzed as specific.

Candidate Repair Solutions and Other-Corrections

Other-initiations of repair that include possible solutions to the trouble differ in whether the solution is offered as a candidate replacement of the trouble source or asserted as a correction of the trouble source. These two alternatives are illustrated in Extract (1) and Extract (2), respectively.

```
(1) Virginia
1 Wes:      °Here you go°
2 Bet:      (But) she[: got so: ma:d.
3 Vir:              [°Thank you°
4           (3.2)
5 Mom: ->   Pam's mother?
6 Bet:      Mm: hm

(2) RCE09
1 Ben:      She's ea(h)ting the Butterwo(h)rth
2           di[e(h)t.
3 Jam: ->     [Bu(h)tterfie(h)ld.
4           (0.9)
5 Ben:      Butterfield.
```

In Extract (1) the other-initiation of repair "Pam's mother?" (line 5) is a candidate solution to the speaker's trouble with the reference "she" (line 2). The candidate is produced with rising intonation, which qualifies the speaker's epistemic stance. The speaker thereby offers this as a possible, but not definitive, solution to the trouble. In contrast, the other-initiation of repair in Extract (2) is produced with falling intonation and an accent on the third syllable, through which the speaker asserts it as a correction of the trouble source. The cases also differ in the epistemic status (Heritage, 2012) of the speakers. In Extract (1), the mother has only indirect knowledge of the event reported in the trouble-source turn, whereas the speaker in Extract (2) has direct knowledge of the correct name.

This practice of other-correction has also been examined by Jefferson (1987) under the rubric of 'exposed correction,' an example of which occurs in Extract (3).

```
(3) Jefferson (1987:87)
1 Pat:      the Black Muslims are certainly more
2           provocative than the Black Muslims
3           ever were.
4 Jo:  ->   The Black Panthers.
5 Pat:      The Black Panthers.
```

To be analyzed as an other-correction, the OIR had to (i) include a possible replacement for the trouble source, (ii) use prosodic resources (an accented syllable and final falling intonation) to assert the replacement as definitive; and (iii) make self-correction (not confirmation) conditionally relevant as a response.

OIRs in and After Next Position

Although the majority of OIRs occur directly after the turn at talk containing the trouble source, a minority of cases occur after this next-turn position (Schegloff, 2000; Wong, 2000; Bolden, 2009; Benjamin, 2012). A distinction between these two positions is crucial for the present investigation because only OIRs in next position to the turn-constructional unit (TCU; Sacks et al., 1974) that contains the trouble-source result in a transition that consist of a gap or overlap, without intervening talk. The OIR in Extract (4) illustrates this point.

```
(4) RCE01 09:56
1 Liz:      I don't- (0.8) I don't know whether to
2           get a maxi dress for my birthday.
3           (0.5)
4 Liz:      I've got one and it['s k- just
```

TABLE 1 | Frequency and proportion of other-initiations of repair in a contiguous next position to trouble-source turn-constructional units (TCUs).

Type	Example	Frequency	Proportion %
Open	*what?, huh?, pardon?, what's that?,* among others	53	31.0
Interrogative words	*who?, when?, where?,* and *what* with falling intonation	11	6.4
Repeats + interrogative word	A: *A plastic bag if you could.* B: *A what?* A: *Plastic bag.*	19	11.1
Full repeats	A: *And we have things to finish.* B: *We have things to finish?* A: *That we started earlier.*	9	5.3
Partial repeats	A: *We could start a little school together.* B: *Little school?* A: *Yeah, like Angel was gonna do.*	20	11.7
Candidate understandings	A: *Nan's birthday on Sunday.* B: *Norms?* A: *No, Shirley.*	42	24.6
Corrections	A: *Transforming Investments* B: *Translating Investments. Sorry.* A: *Translating Investments.*	12	7.0
Other	*I don't know who that person is, who's the guy,* among others	5	2.9

```
5 Cha:  ->              [What's maxi.=Long?
6 Liz:     Really long, yeah.
```

The other-initiation "What's maxi.=Long?" (line 5) does not occur directly after the TCU that contains the trouble source "maxi" (line 2), but rather occurs after the speaker of the trouble source extends her turn with an additional TCU (line 4). The duration of time between the end of the trouble-source TCU and the beginning of the OIR does not constitute a inter-turn gap because it includes intervening talk which affects the timing of the OIR.

All cases of OIR were therefore analyzed for position, following Schegloff (2000), with the requirement that the OIR be in a contiguous next position to the TCU containing the trouble source. The boundaries of TCUs were identified as points of syntactic, intonational, and pragmatic completion (Ford and Thompson, 1996). A total of 51 cases (23% of all cases) did not occur in next position under this definition and were excluded from the analysis of timing.

Identification of Polar Questions

In addition to OIRs, responses to polar questions were also identified in the corpus for comparison. For each recording, a number of polar questions equal to the number of other-initiations was identified, starting at the beginning of the recording. For example, if 10 other-initiations were identified in a recording, the first 10 polar question sequences were then taken from the same recording. Polar questions were defined functionally to include both syntactic questions (i.e., those with verb inversion) and epistemic questions (i.e., statements about information in the recipient's epistemic domain, so-called B-event statements; Labov and Fanshel, 1977).

Measurements and Statistics

The duration of turn transitions were measured from the end of the TCU containing the trouble source to the beginning of the other-initiation of repair, excluding audible inbreaths. Measurements were made manually in ELAN 4.3.3 (Wittenburg et al., 2006) by listening to the audio recording and inspecting the waveform. Two extreme outliers with gap durations greater than 3000 ms were excluded from the quantitative analysis, resulting in a final set of 169 cases. Because the distributions of gap durations were found to deviate substantially from a normal distribution (with skewness and kurtosis values of more than twice their respective standard errors), nonparametric significance tests were used. All statistical tests were performed in R 2.14.0 (R Development Core Team, 2013) with the wilcox_test() function in the coin package (Hothorn et al., 2006).

Results

The Timing of Other-Initiations of Repair

The timing of an other-initiation of repair is the duration of the transition space, measured in milliseconds, between the end of the trouble-source TCU and beginning of the OIR (see Materials and Methods). In Extract (5), the OIR at line 4 occurs after a gap of 514 ms. Hereafter the transition spaces before OIRs are reported in milliseconds in all transcripts, whereas others are given as 10ths of seconds, the standard convention in conversation analysis.

```
(5) Virginia
1 Bet:     They said that Phillips got um (0.5) knee:
2          wa:lking dru::nk at the reception.
3       -> (514 ms)
4 Mom:     Who:?
5 Bet:     Phillips,
```

The density plots in this section present the durations of the transitions between trouble-source TCUs and OIRs along the *x*-axis. Positive values constitute gaps and negative values are overlaps. The density curves represents estimates of the frequency of cases with a given transition time. The peak of the curve corresponds to an estimate of the mode of the distribution.

A comparison between the timing of OIRs and responses to polar questions is presented first (see OIRs and Responses to Polar Questions), after which two comparisons within OIR types are presented: open versus specific (see Open and Specific OIRs) and corrections versus non-corrections (see Corrections and Non-Corrections).

OIRs and Responses to Polar Questions

Figure 1 presents a density plot of the gap durations for OIRs ($n = 169$) and responses to polar questions (PRs; $n = 169$). An inspection of the two distributions reveals that OIRs tend to occur after significantly longer gaps that PRs. The mode gap duration for OIRs is ∼700 ms, whereas the mode for PRs is roughly 300 ms. A Wilcoxon Mann–Whitney Rank Sum Test confirms that the distribution of OIRs tends to have larger values than that of PRs ($Z = 6.5228$, $p < 0.001$, $r = 0.5$). Additional descriptive statistics are given in **Table 2**.

The analysis of the timing of OIRs in general supports the conclusion that OIRs systematically occur after long gaps. The most frequent OIRs do not occur within the same timing window the most frequent PRs, between 0 and 500 ms, but few OIRs occur after 1500 ms. Indeed, if one assumes that the timing of PRs serves as a good proxy for a normal turn transition, as has been argued (Stivers et al., 2009), then the analysis suggests that in this data OIRs typically occur after 400–500 ms of delay beyond the 300 ms duration of a normal transition space.

Open and Specific OIRs

The density plot in **Figure 2** shows the distributions of gap durations for open ($n = 53$) and specific OIRs ($n = 116$) in next position. The density curves indicate that the most frequent gap duration for open OIRs is between 700 and 800 ms, in contrast to approximately 400 ms for specific OIRs. A Wilcoxon Mann–Whitney Rank Sum Test indicates that the two distributions differ significantly ($Z = 1.97$, $p < 0.05$, $r = 0.15$). These results suggest that on average speakers of open OIRs delay ∼300–400 ms more than speakers of specific OIRs.

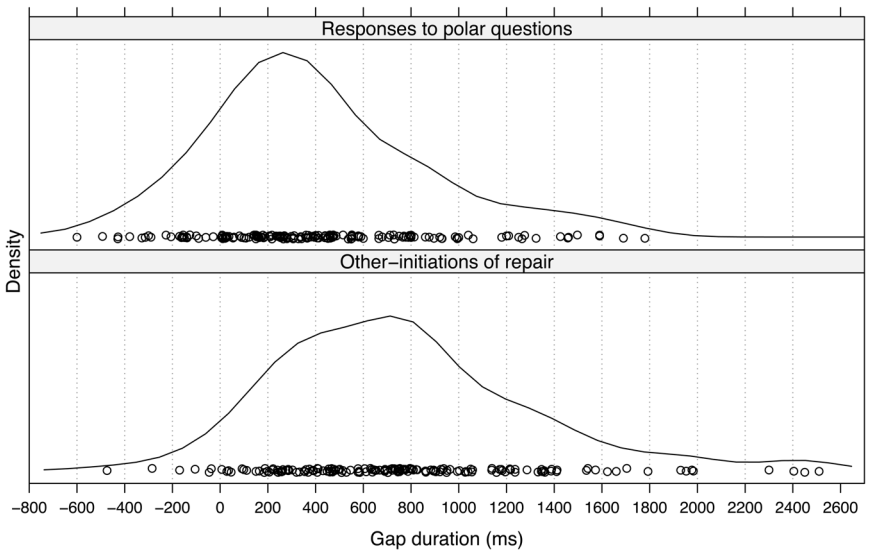


FIGURE 1 | Gap durations (in milliseconds) for other-initiations of repair (OIRs) and responses to polar questions.

TABLE 2 | Descriptive statistics (mean, standard deviation, and median in milliseconds) for gap durations by type.

	Mean (SD)	Median	*N*
Responses to polar questions	397 (475)	339	169
All OIRs	760 (532)	721	169
Specific OIRs	726 (568)	633	116
Open OIRs	835 (439)	787	53
Corrections	412 (422)	274	12
Non-corrections	787 (530)	744	157

Corrections and Non-Corrections

The frequency of other-corrections in the corpus was low, with only 14 cases in total and only 12 cases in next position to the trouble-source TCU (see Materials and Methods). **Figure 3** presents a density plot for the gap durations of other-corrections ($n = 12$) and all other OIRs ($n = 157$) in next position. An inspection of the density plot reveals that, although the number of cases in the collection is small, other-corrections tend to occur earlier than OIRs in general. The most frequent gap duration for other-corrections is between 200 and 300 ms whereas other OIRs most frequently occur after ∼700–800 ms. A Wilcoxon Mann–Whitney Rank Sum Test indicates that other-corrections tend to have shorter gap durations ($Z = -2.64$, $p < 0.01$, $r = 0.20$). These results suggest that in contrast to the bulk of other-initiations, other-corrections do not occur after significant delay and in fact occur within a similar temporal window as responses to polar questions.

What do Speakers Use the Transition Spaces Before OIRs to Accomplish?

The analysis of the timing of OIRs revealed that the most frequent cases occur after gaps of ∼700 ms, in contrast to 300 ms for responses to polar questions. This observation raises the question of what participants use this extra time to accomplish. A qualitative analysis of the transition spaces before OIRs points to three possible answers. The speaker of an OIR, before its production, can:

(1) perform a search for late recognition of the trouble-source turn,
(2) provide an opportunity for the speaker of the trouble-source turn to self-initiate repair, and
(3) produce visual signals, such as facial gestures, that display a lack of recognition and thereby occasion—if not initiate—self-repair.

As the analyses in this section make clear, these possibilities are not necessarily mutually exclusive.

Searching for Late Recognition

In order to produce a relevant next turn, a next speaker must hear, and understand the current turn. The fact that next turns frequently take only 100–300 ms to initiate suggests that the procedure next speakers engage in to recognize the current turn's meaning and action typically occurs quickly, enabling a minimization of gaps between turns (Levinson, 2013). But recognition does not always occur so quickly and can in fact come late, even after a next speaker displays a lack of recognition. This can be seen in the following cases, in which the recognition of a prior turn occurs *after* the next speaker initiates repair.

```
(6) RCE06
1 Alex:     Did you like buy some lemonade earlier,
2           (937 ms)
3 Rob:  ->  Buy some what,=<lemonade,
4 Alex:     (Yeah)
5 Rob:      Yeah yeah yeah.
```

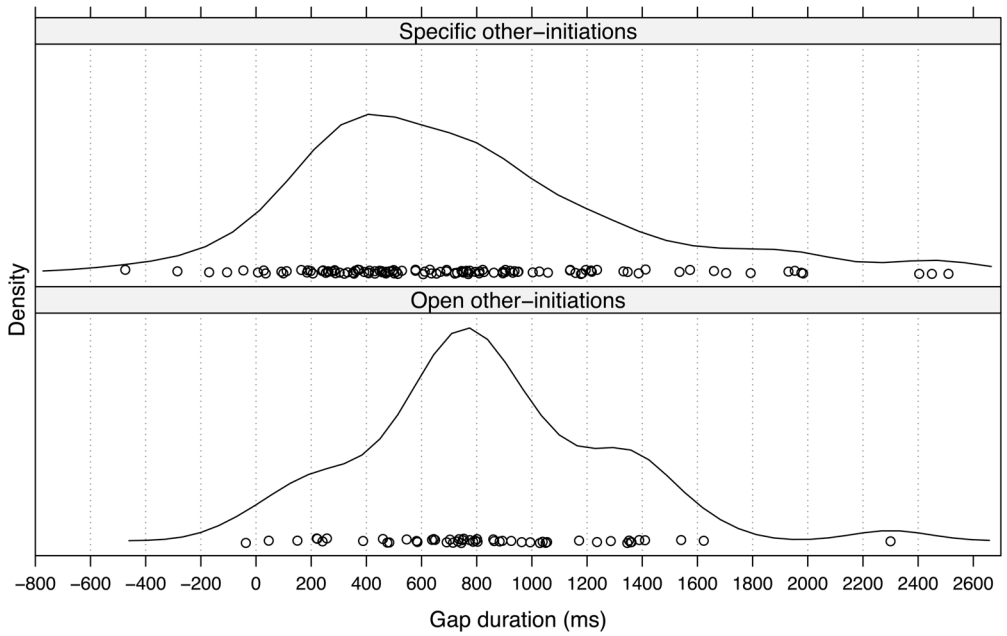


FIGURE 2 | Gap durations (in milliseconds) for open and specific OIRs.

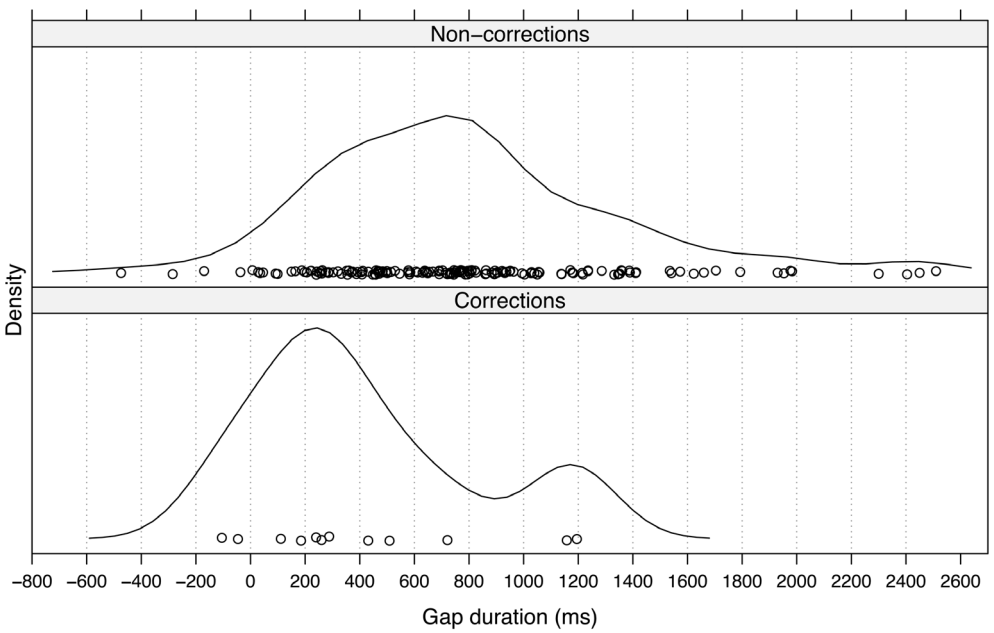


FIGURE 3 | Gap durations (in milliseconds) for corrections and all other OIRs (non-corrections).

(7) Virginia

```
1 Wes:      (Now) you taught 'er howda dance, didn' you?
2           (1236 ms)
3 Vir:   -> Huh? [ (.) [Yeah.
4 Wes:           [Weren['t you teachin' er' some new
5           steps the other day?
6 Vir:      Y:eah.
```

In the first case, Rob apparently fails to hear a word in Alex's question, evidenced by his OIR ("buy some what,"), which locates "lemonade" as a trouble source and makes repetition of this word by Alex conditionally relevant. But before Alex responds, Rob initiates self-repair, producing the very word he claimed, by virtue of his OIR, not to have heard. The word is produced with a prosodic practice known as a left push, noted by the "<" in the transcript,

through which the speaker interdicts the relevance of transition at the completion of a prior TCU (cf. Local and Walker, 2004). Here the left-push is hearable as a 'last millisecond' effort to get the word into the turn before the other responds. In the second case, Virginia apparently fails to hear Wes's question and initiates repair with "Huh?." Then, a moment later, after a micro pause of 140 ms, she answers the question in overlap with Wes's self-repair, claiming in effect to have heard the question, at least well enough to confirm it. These cases suggest that in addition to a procedure that results in 'immediate' recognition of a current turn, next speakers also have available a procedure that can result in 'late' recognition.

Although a search for late recognition is primarily a private process, observable behaviors, such a momentary withdrawal of gaze in the transition space, may reflect this process and thereby render it public. Psycholinguistic research has shown that speakers often avert their gaze when asked questions and that this in turn facilitates remembering and speech planning (Glenberg et al., 1998; Doherty-Sneddon and Phelps, 2005; Markson and Paterson, 2009). Thus the withdrawal of gaze in the transition space before a relevant next turn may be a public exponent of a private search for recognition, as is arguably the case in the next extract. Here, after Heather initiates a new sequence, assessing a taxi driver that she evidently hired the night before, Kelly looks away from Heather in an expanded transition space, apparently engaged in a search for recognition.

(8) RCE28

```
1  Hea:     That taxi driver last night was really
2           friendly.
3       ->  (1346 ms) ((see Figure 4))
4  Kel:     What?
5           (0.6)
6  Hea:     My taxi driver was really friendly.
7           (0.4)
8  Kel:     OH, yea[h.
9  Hea:             [yesterday.
10          (1.5)
11 Kel:     I was like she took a while.
```

The reference to the taxi driver in Heather's turn includes a demonstrative ("that taxi driver"), signaling to Kelly that she should be able to recognize the reference (Himmelmann, 1996). At the completion of Heather's turn, Kelly averts her gaze from Heather and holds this position for ∼1100 ms (see **Figure 4**). The timing of Kelly's look away coincides with the recognizable completion of Heather's turn and thereby shows that Kelly has heard the turn, at least well enough to identify a transition relevance place (Sacks et al., 1974; cf. Holler and Kendrick, 2015). The look away also shows that Kelly has begun to act at just the place where an action by her is relevant. In this way, the withdrawal of gaze at a transition relevance place can be seen as preparatory to an incipient response. In this case, however, Kelly does not produce a relevant response, but rather returns her gaze to Heather and initiates repair. One plausible account of this behavior is that the withdrawal of gaze reflects a search for recognition, one that evidently fails. A search for late recognition is thus one possible use that a next speaker can make of an expanded transition space, one which, if successful, may obviate the need for repair.

FIGURE 4 | Kelly looks away at the completion of the trouble-source turn and holds this position for ∼1100 ms (top). She then returns her gaze to Heather for ∼250 ms before she produces the OIR **(bottom).**

Providing an Opportunity for Self-Initiated Repair

An expansion of the transition space, whether the result of a search for recognition or not, has as an interactional affordance the provision of an opportunity for the current speaker to self-initiate repair and thereby potentially resolve the trouble (Schegloff et al., 1977). Indeed, the absence of a response within a normal transition space can occasion a self-initiation of repair by the current speaker, as can be seen in the following extract.

```
(9) Monopoly Boys
1 Rick:     You have th:e thing I need for the mono:poly.
2           (540 ms)
3 Rick: -> Over there. ((points))
4           (438 ms)
5 Rick: -> The reds.
6 Luke:     Oh, yeah.
```

In the course of a game of Monopoly, Rick notices that Luke has "th:e thing" (i.e., a specific property) that he needs in order to have a monopoly in the game (i.e., to own all the properties of a specific color). At the word "thing" Luke can be seen to begin a visual search of his properties (a set of cards on the side of the game board) that continues into the transition space after Rick's turn. Rather than wait for Luke either to resolve the reference himself or to initiate repair (e.g., with "what" or "what thing"), Rich adds an increment to his turn, which, together with a deictic pointing gesture, specifies the area in Luke's visual field he should search and thereby assists him in the resolution of the problematic reference. After this, too, fails to secure recognition from Luke, Rick again initiates self-repair, replacing "th:e thing" with "the reds" (i.e., the red colored properties), a form of indexical repair that speakers can employ to pursue a response (Bolden et al., 2012). With this, Luke is apparently able to resolve the problematic reference and registers this change of state publicly with "oh" (Heritage, 1984) and confirms Rick's noticing with "yeah."

Here, then, the self-initiation of repair by the current speaker is an alternative to the other-initiation of repair by next speaker. Luke uses the transition spaces that emerge in the course of Rick's turn to search for recognition, a search that, in this case, is publicly observable. Before the search comes to an end, either in late recognition or failure (i.e., an other-initiation of repair), Rick uses the transition spaces as opportunities to self-initiate repair. The practices that current and next speakers employ reveal complementary orientations to their accountability for the intelligibility of the current turn (Garfinkel, 1967). Luke does not initiate repair immediately; he withholds other-initiation to first search for recognition independently. In this way, he holds himself accountable for the recognition of the current turn. Likewise, Rick does not wait for Luke to initiate a repair procedure; he self-initiates repair at the first sign that the recipient has failed to recognize his turn (i.e., the expansion of the transition space and the visible search). In so doing, he orients to his accountability for the intelligibility of his own conduct.

If the current speaker passes on the opportunity to self-initiate repair provided by an expansion of the transition space, the necessity to find a resolution of the trouble falls to the next speaker. That is, if the current speaker does not initiate self-repair, the next speaker may resort to OIR, as Rich does in the next extract.

```
(10) Coffee Chat (simplified)
1 Rich:     ((clears throat))
2           (2.0)
3 Rich:     WE[:LL,
4 Tom:        [That's in: building A?
5           (1286) ((see Figure 5))
6 Rich:  -> Pardon?
```

FIGURE 5 | Rich, on the left, looks down during the trouble-source turn (top). He then turns to look at Tom after the trouble-source turn is complete and holds his gaze on Tom for ~800 ms before he produces an OIR **(bottom)**.

```
7 Tom:      What building are you in?
8 Rich:     Yeah: I'm on the second floor A building.
```

After a lapse in the conversation, in which Rich can be seen to inspect his empty coffee cup, a possible warrant to leave the table, Rich produces what can be heard as a preliminary to his departure from the interaction ("WE:LL," at line 4). In overlap with this, Tom poses a question to Rich that continues on the topic of the talk from before the lapse (i.e., who lives in the same building as Rich in a retirement community). During the question, Rich looks down and forward. ~300 ms after Tom's question comes to completion, Rich turns his head to the left to direct his gaze at Tom (see **Figure 5**). He maintains this position, gazing at Tom, for roughly 800 ms before he produces an OIR ("Pardon?", line 7).

While the entire duration of the transition space before the OIR constitutes an opportunity in which Tom *could* self-initiate repair (e.g., by repeating his question), the roughly 800 ms that Rich holds his gaze on Tom arguably constitutes a space in which Tom *should* self-initiate repair. Heath (1984, p. 253) has argued that gaze, as a display of recipiency, is "sequentially implicative" and "declares an interest in having some particular action occur in immediate juxtaposition with the display." In line with this, Rossano (2006) and Stivers and Rossano (2010) have argued that participants use gaze to signal the relevance of a response. The withdrawal of gaze, in contrast, has been associated with an absence of sequential implicativeness, both at possible sequence completion (Rossano, 2012) and in word searches (Goodwin and Goodwin, 1986). Thus although the current speaker has an opportunity to self-initiate repair and although the deployment of gaze by next speaker arguably displays an expectation that the current speaker should act, the current speaker here passes on the opportunity to self-initiate repair. The initiation of the repair procedure then falls to the next speaker.

Producing Visual Signals

An expansion of the transition space before the production of an OIR also provides space for, and can be the result of, a next speaker's production of visual signals and other visible bodily actions. It has been shown that head movements, such as a lateral tilt or forward extension of the head, can serve to occasion self-repair (Seo and Koshik, 2010) and that particular body movements frequently occur in repair sequences (Rasmussen, 2013; Li, 2014; Floyd et al., in press). In addition, facial gestures, like raising or furrowing one's eyebrows, can be preliminaries to verbal OIRs. In the following extract, after Abbie turns to gaze at Maureen, she raises her eyebrows and holds this position for ∼260 ms before she initiates repair with "Hm:?"

```
(11) Game Night
1 Abbie:     Apparently she's a really spiritual person
2            with a lot of spirituality and stuff like
3            ↑this..hh
4 Maureen:   M.A.?
5            (714) ((see Figure 6))
6 Abbie: -> Hm:?
7 Terry:     Mm:hm[:,
8 Maureen:         [Is it M.A.?
9 Abbie:     Mm:.
```

FIGURE 6 | Abbie, in the middle, looks down during the trouble-source turn (top) and then raises her eyebrows and turns to look at Maureen. She holds this position for a beat (∼260 ms) before she produces an OIR **(bottom)**.

The trouble source that Abbie's OIR locates is itself an OIR which locates the reference to "she" at line 1 as a trouble source and offers the initials "M.A." as a candidate replacement. Abbie's conduct in the transition space – directing her gaze to Maureen, raising her eyebrows, and holding for a beat – not only provides an opportunity for the self-initiation of repair but also constitutes a visible and accountable signal, in the form of a facial gesture, that displays a lack of recognition and a state of recipiency. (Note that Abbie's open OIR "hm:?" lacks the "astonished" prosody associated with open OIRs that signal surprise; Selting, 1988.) In this case, the visual signal does not itself elicit a self-repair; the next speaker goes on to produce a verbal OIR. But elsewhere such visual signals can prompt a current speaker to self-repair his or her talk without a verbal OIR.

In the next extract, Heather self-repairs a place reference in her answer to Kelly's question after Kelly produces a facial gesture that displays a lack of recognition. The question that Kelly asks Heather concerns the amount of time that a friend of Heather's has lived in specific regions of England.

```
(12) RCE28
1 Kel:   It's over ten years, that's pretty much all::
2 Hea:   Yeah he went to my infant school and he went
3        to:: St. Jo:se:ph which is like the school in
4        hhh (0.8) well I think it- (.) counts as
5        Merrow.
```

```
6      -> (1350 ms) ((see Figure 7))
7 Hea:    or like Guildford.=But it's [still Surrey.
8         (0.2)                       [((Kelly nods))
9 Kel:    (Okay)
```

After some initial trouble with the place reference, Heather offers "Merrow" as the name of a place where the friend has lived. In a position in which acceptance of the answer is relevant (e.g., a sequence-closing third, see Schegloff, 2007), a gap of ∼840 ms emerges, at which point Kelly produces a facial gesture – raising her eyebrows and pulling down the corners of her mouth – displaying a lack of recognition (see **Figure 7**). Kelly holds this facial gesture for ∼510 ms, until Heather self-repairs the place reference from "Merrow" to "Guildford," a nearby town. The reference to "Guildford" is apparently recognizable to Kelly, who begins to nod shortly thereafter, and after Heather offers "Surrey" as the name of the county where both places reside, Kelly accepts the answer as adequate and brings the sequence to a close (line 9). This case demonstrates that a facial gesture by a possible next speaker can be sufficient to occasion self-repair. The production of a visual signal within the transition space before an OIR, such as in Extract (11), could therefore be a practice for the resolution of a trouble, one that may obviate the need for a verbal OIR[1].

FIGURE 7 | After Heather's answer to Kelly's question is complete, Kelly gazes at Heather and holds this position for ∼840 ms (top). She then raises her eyebrows, pulls down the corners of her mouth, and holds this facial gesture for ∼510 ms **(bottom)**, until Heather produces a self-repair.

Discussion

The Motivations for and Consequences of Delay Before Other-Initiations of Repair

The quantitative analysis of the timing of OIRs in conversation confirms the observation that OIRs tend to occur after expanded transition spaces. Indeed, if one assumes that the timing of responses to polar questions can serve as a proxy for a normal turn transition, as others have done (Stivers et al., 2009), then the results suggest that other-initiations typically occur after 400–500 ms of delay beyond the 300 ms duration of a normal transition space. But why should this be so? The explanation put forward by Schegloff et al. (1977), discussed previously, is that next speakers who encounter troubles of speaking, hearing, or understanding regularly *withhold* OIR to provide an opportunity for self-initiations of repair. The nature of this explanation is unclear, however. Is this an explanation of a personal motivation for the delay? That is, does a next speaker withhold an OIR *in order to* create an opportunity for the current speaker to resolve the trouble? Or is this an explanation of a public consequence of the delay, one that leaves the question of motivation unanswered? In principle, a delay before an OIR is an opportunity for a self-initiation of repair, whatever its cause.

The analysis of what speakers use the transition spaces before OIRs to accomplish suggests that providing an opportunity for self-initiation of repair does not exhaust the set of possible motives for delay. Although in some cases one can argue that such a motive may lie behind the observed delay (see Providing an Opportunity for Self-Initiated Repair), others point to alternative explanations. The fact that next speakers who initiate repair do, on occasion, recover all, or part of the trouble source after an OIR demonstrates that 'late' recognition is possible. Together with observations of subtle visible bodily actions, such as gaze aversion in an expanded transition space, the qualitative evidence suggests that next speakers who fail to hear or understand the trouble-source turn can engage in a search for this late recognition (see

[1] An examination of visual signals that precede open OIRs – which have the longest transition spaces on average and are therefore the most pertinent to the question – revealed relevant visible bodily actions such as those described in this section in 26.4% of cases ($n = 14$). However, the corpus is not well suited to a quantitative analysis of visible bodily actions. In some cases, the relevant participant is off-camera or his or her face cannot be seen due to the angle of the camera, and in many cases, the participant's eyes cannot be seen well enough to measure his or her gaze. A corpus specifically designed for the temporal analysis of gaze and gesture (e.g., Holler and Kendrick, 2015) is needed to analyze the visual signals that precede OIRs quantitatively.

Searching for Late Recognition). The motivation for the delay in such cases is not to provide an opportunity for self-initiation of repair, but more proximally to resolve the trouble independently, without recourse to an OIR that exposes the trouble and stops the progressivity of the sequence. The evidence also suggests that the visual signals such as facial gestures can precede verbal OIRs. These and other visible bodily actions, including reorientations of the head and torso, take time to produce and can cause a delay – measured in milliseconds of silence – before an OIR. Here, too, the motivation for the delay does not directly concern the self-initiation of repair.

But regardless of the motivation, the consequence is the same: an expanded transition space before an OIR can be a covert signal of trouble and can provide an opportunity for the current speaker to self-initiate repair. Although one can interpret Schegloff et al.'s (1977) explanation as an account of a personal motive, their commitment to uncovering abstract properties and principles of interactional systems suggests that their target was not the individual and his or her motives, but rather an orderliness that transcends such personal concerns. While the results of the investigation are compatible with their explanation, research on the timing of the self-initiation of repair, in particular so-called transition space repairs, is necessary to confirm it. The model that Schegloff et al. (1977). propose predicts that transition space repairs should occur earlier in the transition space than other-initiations. If so, it would provide evidence for a system in which the temporal window for self-initiation precedes that for other-initiation.

Although the results of the investigation are of primary relevance to models of turn-taking and repair in conversation analysis, they may also be of interest to psycholinguists, for whom the timing of turn-taking presents a puzzle. Given that experimental research has shown that speakers need at least 600 ms to plan even a simple word (Indefrey and Levelt, 2004; Indefrey, 2011), processes of language production in conversation must begin well before the current turn ends (Levinson, 2013). Psycholinguists have thus begun to investigate the cognitive processes that enable the minimization of gaps between turns (Magyari and de Ruiter, 2012; Magyari et al., 2014; Sjerps and Meyer, 2015). The results of the current investigation, however, remind us that estimates of central tendency hide a wealth of complex variation, as a number of factors, such as the type of turns involved, influence the timing of transitions. To psycholinguists, the 700 ms of silence that precedes OIRs might be taken to reflect a cognitive process – comprehension gone awry. Indeed, the search for late recognition, as I have called it, may be just such a process. But in face-to-face conversation, the core ecological niche of language (Schegloff, 2006), the line between cognitive processes and socio-interactional ones is blurred. Visible bodily actions, such as an aversion of gaze or a facial gesture, can render otherwise private processes public, at which point they may feed into socio-interactional ones. Even timing alone – a recognition that a speaker has not produced a turn when it was due – can occasion actions such as self-repair. In this way, the private and the public are woven together in an interactional system, and it is within such a system that the silence that precedes OIRs must be understood.

The Preferences for Self-Correction and Specificity

The properties and principles of the repair system, Schegloff et al. (1977) argued, maximize opportunities of self-initiated repair, which come early, and often, and minimize opportunities for other-initiated repair, which as we have seen tend to come late in the transition space. This institutionalized bias in the repair system is known as the preference for self-correction. As evidence of this, Schegloff et al. (1977, p. 379). claimed that other-corrections exhibit special marking and special positioning (e.g., the qualification of epistemic stance or delay within a turn or sequence) that orient to a dispreferred status. With respect to the position of other-corrections within a turn, however, the current investigation finds no evidence for an orientation to dispreference. The results of the analysis, while based on a small sample, show that the other-corrections in the corpus tend to occur without delay, most frequently after 200–300 ms. This suggests that speakers do not withhold other-corrections to provide an opportunity for self-correction. Moreover, the claim that other-corrections typically include qualifications or modulations of epistemic stance has also recently been called into question (Haakana and Kurhila, 2009). Taken together, these findings cast doubt on the status of other-correction as a dispreferred action and suggest that further investigation, based on a larger sample of cases, is warranted.

The relevance of these results to the preference for self-correction itself is less clear. Other-corrections are relatively rare. The entire corpus contains 222 other-initiations, including those that occur after next position; only 6% ($n = 14$) of these are other-corrections. This suggests that many opportunities that speakers may have had to issue a correction simply were not taken. Moreover, other-corrections appear to be restricted to specific types of trouble sources. Of the 14 cases of other-correction, nine locate proper names, or numbers as trouble sources and three target mispronunciations or malapropisms (e.g., "antioxidities" rather than antioxidants). In contrast, other practices for other-initiation do not appear to be restricted in this way. Thus although other-corrections may not be constructed as dispreferred actions (i.e., with delay or qualification), a restriction of other-corrections to specific contexts may nonetheless be evidence of a systematic bias against their use.

In addition to the preference for self-correction, Schegloff et al. (1977) also argue for a preference for specificity in the selection of OIR practices, such that more specific (or 'stronger') other-initiations are preferred over less specific (or 'weaker') ones (cf. Clark and Schaefer, 1987). Two pieces of evidence are given to support this claim. First, if an other-initiation is subject to self-repair within the same turn, the self-repair occurs from a less to a more specific format, but not the inverse. Second, if more than one other-initiation is needed to resolve the trouble, speakers use increasingly specific practices. The current investigation adds two additional pieces of evidence for a preference for specificity. Third, specific other-initiations are more frequent than open other-initiations (only 31% are open, see **Table 2**; cf. Kendrick, in press). And fourth, specific other-initiations tend to occur earlier in the transition space than open other-initiations, in line with the tendency for dispreferred alternatives to be delayed.

The observation that some types of OIRs occur after less delay than others also opens up new avenues for future research. Within the diversity of specific OIRs, for example, one may discover systematic variation. The precise timing of an OIR could indicate a particular epistemic stance, such as whether the OIR signals a trouble of hearing or understanding *per se*, or whether it displays a speaker's surprise or disbelief (see Identification of OIR).

The Intersection of Turn-Taking and Repair

The model of turn-taking that Sacks et al. (1974) proposed accounts for the minimization of gaps in conversation through a set of rules and constraints that motivate fast transitions between turns. Given that OIRs are themselves turns at talk, the observation that OIRs tend to occur after relatively long gaps would therefore appear to undermine this model. In this section – an exploration of the intersection of turn-taking and repair – I first outline a series of systemic constraints on the timing of next turns and then argue that OIRs supersede them, an argument first made by Sacks et al. (1974) and Schegloff et al. (1977), but only partially articulated in their work.

To begin, consider the initial boundary of the transition space. The turn-taking system includes a constraint against more than one speaker at a time, and while more than one speaker at a time is common, it is an unstable state, one which quickly resolves back to a single speaker (Schegloff, 2000). This accounts for the observation that next turns tend to start up at or near possible completions of prior turns, where transition can occur without (or with minimal) violation of the constraint, not sooner.

At the final boundary of the transition space, there are at least three constraints in operation, two of which are rooted in the rules for turn allocation, which provide a motivation for fast transitions (Sacks et al., 1974). Roughly, if no one has been selected to speak next (e.g., by an addressed question), a speaker may self-select to take a turn. If more than one speaker self-selects, the first to start has rights to the turn and the second starter should cede the turn to the first. These rules establish a motivation for next speakers to start up early and thereby minimizes the gaps between turns.

If no one self-selects, however, the current speaker may continue his or her turn. The possibility that such a continuation may be imminent also provides for the minimization of gaps, as next speakers aim to begin before this occurs. Although the time course of this rule is unknown, a computational corpus study of Dutch telephone conversations by Bosch et al. (2005) provides a useful estimate. The duration of silences between utterances within a turn was found to be greater than the duration of silences between turns, with mean durations of 520 and 380 ms, respectively. This is compatible with a model of turn-taking in which an opportunity for self-selection by next speaker temporally precedes that for continuation by current speaker. The expanded transition spaces before OIRs are therefore the result of a connivance: the next speaker passes the opportunity (or obligation, in the case of current-selects-next) to speak, and the current speaker, the one who produced the trouble source, passes the opportunity to continue the turn. At ∼700 ms, the average other-initiation of repair occurs after the absence of a continuation by the current speaker would be recognizable. Indeed, there is evidence that 700 ms may in fact be a generic threshold in conversation (Kendrick and Torreira, 2012, 2014; Roberts and Francis, 2013), perhaps for this very reason.

A third constraint at the final boundary of the transition space is grounded not in the rules for turn allocation but in the potential for silences in conversation to become meaningful, as signals of interactional trouble (Jefferson, 1986, 1983a; Pomerantz, 1984; Schegloff, 1988; inter alia). In an adjacency-pair sequence, to cite but one context, even a slight delay beyond a normal transition increases the likelihood that the second pair part will have a dispreferred turn format, and a long delay, on the order of 700–800 ms, is a reliable signal that a dispreferred response is imminent (Kendrick and Torreira, 2014). The semiotics of silence is therefore an additional basis for a constraint on the timing of next turns, one that, like the rules for turn allocation, creates a bias toward fast transitions and the attendant minimization of gaps.

Given the existence of systemic constrains on the timing of next turns and the observation that OIR occur after significantly longer gaps than other turns, one solution to this apparent puzzle naturally presents itself: OIRs may trump the rules of the turn-taking system. Indeed, this appears to be the tack taken, though only partially articulated, by Sacks et al. (1974) and Schegloff et al. (1977). The timing of OIR, they argued, reveals "the independent status of the repair organization, whose operation may supersede otherwise operative aspects of the turn-taking organization" (Schegloff et al., 1977, p. 374). Although they do not elaborate this point, they do provide one additional example. In a discussion of second-starter supersession (i.e., methods whereby a second speaker to self-select may win the turn), Sacks et al. (1974, p. 720) observe that "when a self-selector's turn-beginning reveals his turn's talk to be prospectively addressed to a problem of understanding [a] prior utterance, he may by virtue of that get the turn, even though at the turn-transfer another started before him." In other words, a second speaker to self-select takes priority if he or she produces an other-initiation of repair. This, then, is evidence that participants in conversation orient to resolving troubles of hearing and understanding as a "priority activity" (Sacks et al., 1974, p. 720), one which takes precedence over rules of the turn-taking system that motivate fast turn transitions.

A consequence of the priority of OIR, one which to my knowledge has not previously been registered, is that OIRs may freely start up in overlap with a post-trouble-source turn or TCU, and need not employ the practices for overlap competition described by Schegloff (2000), such as increases in volume, speech rate, or pitch. In each of the cases below, an OIR starts up in overlap with a turn or TCU that intervenes between it and the trouble-source TCU. Although the OIR is not designed as competitive, in each case the speaker of the prior turn or TCU drops out, ceding the turn to the speaker of the OIR.

```
(13) KC-4:2 (Sacks et al., 1974, p. 720)
1 R:        Hey::, the place looks different.
2 F:        Yea::hh.
3 K:  ->    Ya have to see ou[r new-
4 D:  ->                     [It does?
5 R:        Oh yeah
```

```
(14) WG 4-13-nh (Benjamin, 2013, p. 188)
1 Hal:      he may be victimized on it
2           (0.6)
3 Hal: ->   I'm not sure he's ma[king]
4 Nix: ->                       [you ] mean by his lawyer
5 Nix:      [hhhhhhhhh]
6 Hal:      [yeah or] (.) somebody else
```

```
(15) CallHome 6079 (Benjamin, 2013, p. 119)
1 A:        it was [so: nice] it was so nice =
2 B:               [hhhhh @ ]
3 A:        = that they came I can't even tell you
4           like.hhhh like (0.6) just seeing them
5           like I was performing to them
6           (.)
7     ->    like I was sm[iling at th]em like
8 B:  ->                [to who. ]
9 A:        .hhhh Juliette, Sam: and (.) and Tara
```

The fact that OIRs win the turn without the need for competition and, moreover, that trouble-source speakers respond to OIRs with no delay is further evidence that repair is a priority activity in conversation (see also Extract 4).

The data above also illustrate yet another intersection between turn-taking and repair. As noted previously, the imminent possibility that a current speaker may continue his or her turn if no one self-selects builds a motivation for fast transitions into the turn-taking system. Note, however, that the OIRs in Extracts (14) and (15) occur in overlap with a continuation by the current speaker (at the first arrowed lines). This demonstrates that OIRs supersede not only first-starters in self-selection, but also continuations by current speakers. The window of opportunity for OIRs is thus larger than for other next turns, which are subject to constraints on turn allocation and overlap that OIRs appear to out rank.

This is not to say that the timing of OIR is without constraint. The organization of repair imposes certain constraints on the timing of OIRs vis-à-vis the selection of OIR practices (see Robinson, 2014). An open OIR locates a trouble-source TCU exclusively via adjacency and is therefore positionally restricted. If a next speaker fails to hear or understand a TCU well enough to employ a specific OIR practice, then the window of opportunity to use an open OIR has an outer bound: the possible completion of a subsequent TCU that intervenes between the trouble-source TCU and the open OIR. In other words, the opportunity space for open OIR is a one-TCU interval (Robinson, 2014). But does *this* constraint, which operates for open OIRs, establish a motivation for fast turn transitions? It provides an outer bound for the timing of open OIRs, but given that an open OIR can in principle occur in overlap with a subsequent TCU (before its possible completion) and take priority, it would not systematically motivate a fast transition on the order of 100–300 ms between the trouble-source TCU and the OIR.

What, then, of the potential for silences to be meaningful, as signals of interactional trouble? Might this provide a motivation for fast transitions? The answer becomes clear once one registers that OIRs are themselves signals of interactional trouble. Although a covert signal of trouble like silence will be at cross-purposes with some incipient actions (e.g., agreement), it can also point in the same direction as an incipient action (e.g., rejection), in which case it is interactionally advantageous. An expanded transition space can indicate that a next speaker has encountered trouble, but it also provides an opportunity for the resolution of the trouble. There is reason to believe that speakers do not *avoid* this covert signal (e.g., by initiating repair quickly), but rather *exploit* it. As a motivation for fast turn transfers, the potential for silences in conversation to be meaningful signals of interactional trouble is thus context-sensitive, operating only for next turns not themselves designed as signals of trouble.

In sum, the rules and constraints of the turn-taking system that motivate fast transitions – concerning overlap management, turn allocation, and the semiotics of silence – neither rule nor constrain the timing of OIRs. The priority given, at the intersection of turn-taking and repair, to the resolution of troubles in hearing and understanding provides a systemic explanation for the observation that transitions before OIRs tend to be longer than those before other next turns, for which the rules and constraints of the turn-taking system remain operative.

Acknowledgments

I would like to thank Jeff Robinson, Trevor Benjamin, and Steve Levinson for useful comments on earlier versions of this article, as well as Francisco Torreira, from whom I have learned a great deal. This research was made possible by the financial support of the Language and Cognition Department at the Max Planck Institute for Psycholinguistics.

References

Benjamin, T. (2012). When problems pass us by: using 'You Mean' to help locate the source of trouble. *Res. Lang. Soc. Interact.* 45, 82–109. doi: 10.1080/08351813.2012.646742

Benjamin, T. (2013). *Signaling Trouble: On the Linguistic Design of Other-Initiation of Repair in English Conversation*. Ph.D. Dissertation, University of Groningen, The Netherlands.

Benjamin, T., and Walker, T. (2013). Managing problems of acceptability through high rise-fall repetitions. *Discourse Process.* 50, 107–138. doi: 10.1080/0163853X.2012.739143

Bolden, G. B. (2009). "Implementing delayed actions," in *Conversation Analysis: Comparative Perspectives*, ed. J. Sidnell (Cambridge: Cambridge University Press), 326–353. doi: 10.1017/CBO9780511635670.012

Bolden, G. B., Mandelbaum, J., and Wilkinson, S. (2012). Pursuing a response by repairing an indexical reference. *Res. Lang. Soc. Interact.* 45, 137–155. doi: 10.1080/08351813.2012.673380

Bosch, L.-T., Oostdijk, N., and Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Commun.* 47, 80–86. doi: 10.1016/j.specom.2005.05.009

Clark, H. H., and Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Lang. Cogn. Process.* 2, 19–41. doi: 10.1080/01690968708406350

Doherty-Sneddon, G., and Phelps, F. G. (2005). Gaze aversion: a response to cognitive or social difficulty? *Mem. Cogn.* 33, 727–733. doi: 10.3758/BF03195338

Drew, P. (1997). 'Open' class repair initiators in response to sequential sources of troubles in conversation. *J. Pragmat.* 28, 69–101. doi: 10.1016/S0378-2166(97)89759-7

Egbert, M., Golato, A., and Robinson, J. D. (2009). "Repairing reference," in *Conversation Analysis: Comparative Perspectives*, ed. J. Sidnell (Cambridge: Cambridge University Press), 104–132. doi: 10.1017/CBO97805116356 70.005

Floyd, S., Manrique, E., Rossi, G., and Torreira, F. (in press). The timing of visual bodily behavior in repair sequences: evidence from three languages. *Discourse Process*. doi: 10.1080/0163853X.2014.992680

Ford, C. E., and Thompson, S. A. (1996). "Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns," in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 134–184.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Glenberg, A. M., Schroeder, J. L., and Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Mem. Cogn.* 26, 651–658. doi: 10.3758/BF03211385

Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociol. Inq.* 50, 272–302. doi: 10.1111/j.1475-682X.1980.tb00023.x

Goodwin, C. (1981). *Conversational Organization: Interaction between Speakers and Hearers*. New York: Academic Press.

Goodwin, M. H., and Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica* 62, 51–76.

Haakana, M., and Kurhila, S. (2009). "Other-correction in everyday interaction: some comparative aspects," in *Talk in interaction: Comparative Dimensions*, eds M. Haakana, M. Laakso, and J. Lindström (Helsinki: Finnish Literature Society), 152–179.

Heath, C. (1984). "Talk and recipiency: sequential organization in speech and body movement," in *Structures of Social Action: Studies in Conversation Analysis*, eds J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 247–265.

Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *J. Phon.* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002

Hepburn, A., and Bolden, G. B. (2013). "The conversation analytic approach to transcription," in *The Handbook of Conversation Analysis*, eds J. Sidnell and T. Stivers (Malden, MA: Blackwell Publishing Ltd).

Heritage, J. (1984). "A change-of-state token and aspects of its sequential placement," in *Structures of Social Action: Studies in Conversation Analysis*, eds J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 299–345.

Heritage, J. (2012). Epistemics in action: action formation and territories of knowledge. *Res. Lang. Soc. Interact.* 45, 1–29. doi: 10.1080/08351813.2012. 646684

Himmelmann, N. P. (1996). "Demonstratives in narrative discourse: a taxonomy of universal uses," in *Studies in Anaphora*, ed. B. A. Fox (Amsterdam: John Benjamins Publishing Company), 205–254.

Holler, J., and Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. *Front. Lang. Sci.* 6:98. doi: 10.3389/fpsyg.2015.00098

Hothorn, T., Hornik, K., van de Wiel, M. A., and Zeileis, A. (2006). A Lego system for conditional inference. *Am. Stat.* 60, 257–263. doi: 10.1198/000313006X118430

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2:255. doi: 10.3389/fpsyg.2011. 00255

Indefrey, P., and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001

Jefferson, G. (1973). A case of precision timing in ordinary conversation: overlapped tag-positioned address terms in closing sequences. *Semiotica* 9, 47–96. doi: 10.1515/semi.1973.9.1.47

Jefferson, G. (1983a). Notes on a possible metric which provides for a 'Standard Maximum' silence of approximately one second in conversation. *Tilburg Papers Lang. Lit.* 42, 1–83.

Jefferson, G. (1983b). On a failed hypothesis: 'Conjunctionals' as overlap-vulnerable. *Tilburg Papers Lang. Lit.* 28, 1–33.

Jefferson, G. (1984). "Notes on some orderlinesses of overlap onset," in *Discourse Analysis and Natural Rhetoric*, eds V. D'Urso and P. Leonardi (Padua: Cleup Editore), 11–38.

Jefferson, G. (1986). Notes on 'Latency' in overlap onset. *Hum. Stud.* 9, 153–183. doi: 10.1007/BF00148125

Jefferson, G. (1987). "On exposed and embedded correction in conversation," in *Talk and Social Organization*, eds G. Button and J. R. E. Lee (Clevedon: Multilingual Matters), 86–100.

Kendrick, K. H. (in press). Other-initiated repair in English. *Open Linguist.* 1, 164–190. doi: 10.2478/opli-2014-0009

Kendrick, K. H., and Torreira, F. (2012). The timing and construction of preference: a quantitative study. *Paper presented at the Discourse, Communication, and Conversation Conference*, Loughborough University, Loughborough.

Kendrick, K. H., and Torreira, F. (2014). The timing and construction of preference: a quantitative study. *Discourse Process.* doi: 10.1080/0163853X.2014.955997

Kitzinger, C. (2013). "Repair," in *The Handbook of Conversation Analysis*, eds J. Sidnell and T. Stivers (Malden: Blackwell Publishing Ltd), 229–256. doi: 10.1080/0163853X.2014.955997

Labov, W., and Fanshel, D. (1977). *Therapeutic Discourse: Psychotherapy as Conversation*. New York: Academic Press.

Lerner, G. H. (2003). Selecting next speaker: the context-sensitive operation of a context-free organization. *Lang. Soc.* 32, 177–201. doi: 10.1017/S004740450332202X

Levinson, S. C. (2013). "Action formation and ascription," in *The Handbook of Conversation Analysis*, eds J. Sidnell and T. Stivers (Malden, MA: Wiley–Blackwell Publishing Ltd), 101–130.

Li, X. (2014). Leaning and recipient intervening questions in Mandarin conversation. *J. Pragmat.* 67, 34–60. doi: 10.1016/j.pragma.2014.03.011

Local, J., and Walker, G. (2004). Abrupt-joins as a resource for the production of multi-unit, multi-action turns. *J. Pragmat.* 36, 1375–1403. doi: 10.1016/j.pragma.2004.04.006

Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., and Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *J. Cogn. Neurosci.* 26, 2530–2539. doi: 10.1162/jocn_a_00673

Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. 10.3389/fpsyg.2012. 00376

Markson, L., and Paterson, K. B. (2009). Effects of Gaze-aversion on visual-spatial imagination. *Br. J. Psychol.* 100, 553–563. doi: 10.1348/000712608X371762

Mondada, L. (2006). Participants' online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence. *Discourse Stud.* 8, 117–129. doi: 10.1177/1461445606059561

Mondada, L. (2007). Multimodal resources for turn-taking pointing and the emergence of possible next speakers. *Discourse Stud.* 9, 194–225. doi: 10.1177/1461445607075346

Pomerantz, A. (1984). "Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes," in *Structures of Social Action: Studies in Conversation Analysis*, eds J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 57–101.

R Development Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.r-project.org/

Rasmussen, G. (2013). Inclined to better understanding—the coordination of talk and 'leaning forward' in doing repair. *J. Pragmat.* 65, 30–45. doi: 10.1016/j.pragma.2013.10.001

Roberts, F., and Francis, A. L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *J. Acoust. Soc. Am.* 133, EL471–EL477. doi: 10.1121/1.4802900

Roberts, F., and Robinson, J. D. (2004). Interobserver agreement on first-stage conversation analytic transcription. *Hum. Commun. Res.* 30, 376–410. doi: 10.1111/j.1468-2958.2004.tb00737.x

Robinson, J. D. (2006). Managing trouble responsibility and relationships during conversational repair. *Commun. Monogr.* 73, 137–161. doi: 10.1080/03637750600581206

Robinson, J. D. (2014). What 'What?' tells us about how conversationalists manage intersubjectivity. *Res. Lang. Soc. Interact.* 47, 109–129. doi: 10.1080/08351813.2014.900214

Robinson, J., and Kevoe-Feldman, K. (2010). Using full repeats to initiate repair on others' questions. *Res. Lang. Soc. Interact.* 43, 232–259. doi: 10.1080/08351813.2010.497990

Rossano, F. (2006). When the eyes meet: using gaze to mobilize response. *Paper Presented at the International Conference on Conversation Analysis*, Helsinki.

Rossano, F. (2012). *Gaze Behavior in Face-to-Face Interaction.* Ph.D. thesis, Radboud University, Nijmegen.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language (Baltim)* 50, 696–735.

Schegloff, E. A. (1987). "Recycled turn beginnings: a precise repair mechanism in conversation's turn-taking oragnisation," in *Talk and Social Organisation*, eds G. Button and J. R. E. Lee (Phiadelphia: Multilingual Matters), 70–93.

Schegloff, E. A. (1988). On an actual virtual servo-mechanism for guessing bad news: a single case conjecture. *Soc. Probl.* 35, 442–457. doi: 10.2307/800596

Schegloff, E. A. (1992). Repair after next turn: the last structurally provided defense of intersubjectivity in conversation. *Am. J. Soc.* 97, 1295–1345. doi: 10.1086/229903

Schegloff, E. A. (1997). Practices and actions: boundary cases of other-initiated repair. *Discourse Process.* 23, 499–545. doi: 10.1080/01638539709545001

Schegloff, E. A. (1998). Body torque. *Soc. Res. (N. Y.)* 65, 535–596.

Schegloff, E. A. (2000). When 'Others' initiate repair. *Appl. Linguist.* 21, 205–243. doi: 10.1093/applin/21.2.205

Schegloff, E. A. (2006). "Interaction: the infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is situated," in *Roots of Human Sociality: Culture, Cognition and Interaction*, eds N. J. Enfield and S. C. Levinson (Oxford: Berg Publishers), 70–96.

Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511791208

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language (Baltim)* 53, 361–382.

Selting, M. (1988). The role of intonation in the organization of repair and problem handling sequences in conversation. *J. Pragmat.* 12, 293–322. doi: 10.1016/0378-2166(88)90035-5

Seo, M.-S., and Koshik, I. (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *J. Pragmat.* 42, 2219–2239. doi: 10.1016/j.pragma.2010.01.021

Sjerps, M. J., and Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition* 136, 304–324. doi: 10.1016/j.cognition.2014.10.008

Stivers, T., Enfielda, N. J., Browna, P., Englertb, C., Hayashic, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Stivers, T., and Rossano, F. (2010). Mobilizing response. *Res. Lang. Soc. Interact.* 43, 3–31. doi: 10.1080/08351810903471258

Strömbergsson, S., Hjalmarsson, A., Edlund, J., and House, D. (2013). "Timing responses to question S in dialogue," in *Proceedings of Interspeech 2013*, (Lyon: International Speech Communication Association), 2584–2588.

Svennevig, J. (2008). Trying the easiest solution first in other-initiation of repair. *J. Pragmat.* 40, 333–348. doi: 10.1016/j.pragma.2007.11.007

Wilkinson, S., and Kitzinger, C. (2006). Surprise as an interactional achievement: reaction tokens in conversation. *Soc. Psychol. Q.* 69, 150–182. doi: 10.1177/019027250606900203

Wilson, T., and Zimmerman, D. (1986). The structure of silence between turns in two-party conversation. *Discourse Process.* 9, 375–390. doi: 10.1080/01638538609544649

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). *ELAN: A Professional Framework for Multimodality Research (2006)*. Available at: http://tla.mpi.nl/tools/tla-tools/elan/

Wong, J. (2000). Delayed next turn repair initiation in native/non-native speaker English conversation. *Appl. Linguist.* 21, 244–267. doi: 10.1093/applin/21.2.244

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 10 March 2015
doi: 10.3389/fpsyg.2015.00251

Expanded transition spaces: the case of Garrwa

Rod Gardner[1]* *and Ilana Mushin*[2]

[1] School of Education and Professional Studies, Griffith Institute for Educational Research, Griffith University, Mount Gravatt, QLD, Australia, [2] School of Languages and Cultures, University of Queensland, QLD, Australia

Accounts of turn-taking in much of the CA literature have largely focused on talk which progresses with minimal gaps between turns at talk, longer gaps being found to be symptomatic of, for example, engagement in non-talk activities, or as indicators of some kind of trouble in the interaction. In this paper we present an account of turn-taking in conversations between Indigenous Australians where longer gaps are frequent and regular. We show that in sequences of such slow-paced conversation, gaps are not always treated as problematic, nor are they associated with non-talk activities that might inhibit talk. In such contexts we argue that there is less orientation to gap minimization, reflecting a lack of pressure for continuous talk. We also discuss qualitative differences in the nature of the gaps between turns in which there is a selection of next speaker, and those where no next speaker has been selected. Finally we consider whether such talk is a feature of Indigenous Australian conversation, or a more widespread practice.

Keywords: conversation analysis, transition spaces, turn-taking, Aboriginal conversation, conversation and culture

OPEN ACCESS

Edited by:
Judith Holler, Max Planck Institute Psycholinguistics, Netherlands

Reviewed by:
Kobin H. Kendrick, Max Planck Institute for Psycholinguistics, Netherlands
Jack Sidnell, University of Toronto, Canada

***Correspondence:**
Rod Gardner, School of Education and Professional Studies, Griffith University, Mount Gravatt Campus, 176 Messines Ridge Road, Mount Gravatt, QLD 4122, Australia
r.gardner@griffith.edu.au

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *30 October 2014*
Accepted: *19 February 2015*
Published: *10 March 2015*

Citation:
Gardner R and Mushin I (2015) Expanded transition spaces: the case of Garrwa. Front. Psychol. ***6****:251. doi: 10.3389/fpsyg.2015.00251*

Introduction

This report had its genesis in a project investigating Aboriginal Australian conversation in the Garrwa language, with a focus on turn-taking practices (Gardner and Mushin, 2007; Mushin and Gardner, 2009, 2011; Gardner, 2010). One observation that emerged from this study was that the pace of these conversations appeared overall to be slower than has been reported for conversation in most of the literature on turn-taking. A specific feature of this slowness was seen in the distribution of inter-turn gaps of silence—the time it took for a next speaker to begin talking when the conversational floor was free. Ethnographic reports of Australian Aboriginal conversation have suggested a tolerance for greater gaps between turns than for Anglo Australian conversation (Walsh, 1991; Eades, 2000, 2007). This prompted the question whether this perceived slower pace indicated something about the fundamental ways in which Aboriginal Australians conduct their conversations, particularly in the overall pace of the talk and timing of speaker change. One hypothesis for these extended gaps between turns is, as Walsh (1991) implied, that Aboriginal Australians are orienting to a different set of rules for conversational turn-taking than are found in those societies that have been the basis of most

Abbreviations: ABL, ablative; ACC, accusative; ALL, allative; BARRI, a discourse particle; CONJ, conjunction; DAT, dative; DEM, demonstrative; DS, different subject (switch reference marker); ERG, ergative; FUT, future; HAB, habitual; IMP, imperative; INTENS, intensifier; KANYI, verbal morpheme (yet to be labeled); LOC, locative; NA, a discourse particle; NEG, negative particle; PA, past tense; PURP, purposive; WA, grammatical morpheme (yet to be labeled); 1sg, first person singular; 2sg, second person singular; 1duIncl, first person dual inclusive (you and me); 3du, third person dual; 1plncl, first person plural inclusive; 1plExcl, first person plural exclusive; 3pl, third person plural.

investigations of turn-taking phenomena, starting with Sacks et al.'s (1974) seminal paper on the systematics of turn-taking using English language data. Walsh suggested that Aboriginal talk is "broadcast" to all present rather than to specific recipients, and further, that in a "non-dyadic" mode of speaking. He also claimed that Aboriginal speakers have a much greater tolerance for silence in conversation.

An alternative hypothesis is based on what Schegloff (2007) has posited as possible departures from interactional formats familiar to Western industrialized nations (which) involve what might be called "differences in the value of variables." Under this hypothesis, the basic rules for turn-taking are the same, but the lengths of time that count as silence may be calibrated differently across cultures. This suggestion has received support from a study by Stivers et al. (2009), which examined the delays (or "response offsets") in responding to polar questions across 10 languages from five continents that were different in terms of language family and culture. They established that longer delays in response were associated with four parameters: if a question was not answered rather than answered; if a question was disconfirmed rather than confirmed; if a response was verbal rather than visible (such as a head nod); and if the speaker had no gaze contact with the recipient rather than gaze contact. They found a mean response time of just over 0.2 s, but variation across languages, with Japanese the shortest at less than 0.1 s, and Danish the longest at a little under 0.5 s. The overall mode was 0 s (supporting the Sacks, Schegloff, and Jefferson claim reported below that turn transitions with no gap and no overlap are common), with a mode variation between 0.0 and 0.2 s across the 10 languages. They found that in all languages the response time was between zero and half a second, which they argue suggests a universal across all languages for gap minimization. They found no correlations of length of delays between closely related languages nor between similar cultures. The variations that they found in response offset times arose from what they termed "a different cultural "calibration" of delay" (p. 10590), which, they claim, has to do with the general pace of conversations and the general tempo of life in the communities in which these languages are spoken.

What we report in this paper is that that these Garrwa speakers have tolerance for silences between turns that appears to be greater than it is for English conversation, including Anglo-Australian conversation, and indeed there is some evidence (cf. Gardner, 2010) that this appears to be stronger than for any of the 10 languages reported in Stivers et al. (2009). However, we provide evidence that in some Anglo-Australian conversations, under certain situational conditions, there also appears to be a greater tolerance for lengthy interturn silences. In these conversations, couples were at home alone in the evening, engaging in "non-focused" talk (cf. Couper-Kuhlen, 2010). This leads us to the hypothesis that what we are finding is not any difference in the ways in which speakers allocate turns of talk, but rather that the circumstances in which the talk is occurring may be what is leading to "differences in the values of variables," in this case the length of interturn silences. We also find strong evidence that Aboriginal speakers of Garrwa generally adhere to the rules of turn-taking as proposed by Sacks et al. (1974).

Based on these observations, the questions we address in this paper are whether delayed onset of talk by a next speaker:

- constitutes a different system of turn-taking among Garrwa from that outlined in Sacks, Schegloff, and Jefferson;
- can be accounted for within the existing parameters for turn-taking without resorting to cross-cultural explanations.

Sacks et al. (1974) proposed a set of rules that derive from two proposals: first, that a turn at talk is made up of a word, phrase or sentence that can stand alone and make full sense in the context of the conversation (a "turn constructional unit" or TCU)[1], and second, that there is a short period of time at the end of a TCU within which change of speaker is warranted by the rules (the "transition relevance place" or TRP). Where a "normal" transition-relevance place begins and ends is not often discussed explicitly in the literature, but the normal, or default space has been declared to be one "beat of silence," i.e., the time it takes to say a single syllable at normal rate (Jefferson, 1984; Schegloff, 2000). Wells and Macfarlane (1998) suggest that it extends from a final, turn-ending projecting accent in a TCU to the onset of a next speaker's talk, which may typically be about two beats, or 0.2 s[2]. If someone starts speaking outside the TRP, this can be treated by participants as problematic.

There are two rules, which specify how change of speaker occurs. The first of these is divided into three parts. The first of these, rule 1a, states that if in the course of a turn the speaker of that turn selects someone to speak next, for example by naming them, by gaze, by touch or by asking a question that only one other participant has the knowledge to answer. Under such circumstances, the speaker who has been selected is obliged to begin speaking at the TRP that occurs at the end of the TCU that is in progress. The second part, rule 1b, comes into play if the current speaker does not select a next speaker, and states that at the next TRP, any speaker other than the current speaker may start speaking (or may "self-select,") and if there is more than one other participant, the first to start has rights to the floor. The third part, rule 1c, states that if no other speaker self-selects under rule 1b, then the current speaker may (but need not) continue speaking, thereby producing a second TCU in their turn. A second rule is necessary because under rule 1c, there has been no change of speaker, so in order to state how such change occurs, Sacks, Schegloff, and Jefferson made explicit that if current speaker does continue under rule 1c, then the three parts of the first rule are recycled until change of speaker does occur.

This simple set of rules has a power that may not be immediately obvious, namely that, first, the system is built to ensure that speaker change occurs frequently, but also, second, bids for speakership need to be made with precision timing that

[1]Since 1974, the notion of what constitutes a turn-constructional unit has developed further, see for example Schegloff (1996) and Ford and Thompson (1996).

[2]Where a "normal" transition-relevance place begins and ends is not often discussed explicitly in the literature, but the normal, or default space has been declared to be one "beat of silence," i.e., the time it takes to say a single syllable at normal rate (Jefferson, 1984; Schegloff, 2000). Wells and Macfarlane (1998) suggest that it extends from a final, turn-ending projecting accent in a TCU to the onset of a next speaker's talk, which may typically be about two beats, or 0.2 s.

requires preparation for the bid while the current TCU is still underway.

First we demonstrate that the participants in these Garrwa conversations on many occasions allocate turns very much in the way predicted by Sacks et al. (1974). Then we examine some examples of talk with expanded transition spaces, in which gaps between turns appear to be longer that Sacks, Schegloff, and Jefferson talk of as a normal TRP. They claim as one of their "gross observations" that speaker transitions with no gap or overlap are common, and together with transitions with slight gaps or overlaps, they make up the "vast majority of transitions" (p. 701). In the Garrwa conversations, we commonly find a tolerance for silence between turns of up to several seconds. Sometimes these silences can be accounted for by non-talk activities, such as drinking or grooming that may be distracting the speaker from the ongoing talk, but some activities, such as grooming, do not in and of themselves disable the ability to talk. We then examine whether there are differences in the length of long silences following current speaker selection of the next speaker (rule 1a) or following self-selection of next speaker (rule 1b), and we report that we do find regular differences in the length of the long silences after application of rule 1a, compared with turns after application of rule 1b. Finally, we pose the question of the degree to which these unusually long silences are a phenomenon of Aboriginal Australian interactions—and perhaps the interactions of some other indigenous peoples (cf. Scollon and Scollon, 1981; Hoymann, 2010)—or a more general practice of talk-in-interaction that derives from a lack of pressure for continuous talk, associated with situational factors such as intimates just "hanging out" with nothing particular to talk about, in familiar surroundings with no pressure to "get things done." We then begin to examine the extent to which long gaps between turns are a cultural phenomenon by examining some examples of long gaps between turns in two Anglo-Australian English language conversations.

A number of Conversation Analysts have suggested that delayed responses are regularly associated with problems in the talk. There may be talk-internal reasons for the silences, such as word searches (Goodwin and Goodwin, 1987; Hayashi, 2003), attempting to gain the attention of another speaker (Goodwin, 1981), speech impairment of one or more of the participants (Goodwin, 1995b). Other delays may occur prior to dispreferred, or non-agreeing responses (Pomerantz, 1984a), or during resistance to requests (Davidson, 1984)[3]. Pomerantz (1984b) points to recipient problems due to unclear references in the prior speaker's turn, a lack of recipient knowledge that the prior speaker had assumed, and recipient disagreement to account for silences of up to about a second between a first pair part (such as a question) and a second pair part (such as an answer). Davidson (1984) notes that silence after a first pair part can be a result of "puzzlement, or lack of clarity about exactly what's being offered" (p. 127), as well as difficulty in hearing and doubt about the acceptability of the proposal (p. 103). In some forms of institutional talk, long silences may be tolerated, for example in language (and perhaps other) classrooms where a teacher may wait a long time for a student answer to her question (Gardner, 2007). Silences after first pair parts can be precursors to a "potential rejection" (p. 103). As Jefferson (1986) puts it, 'for the vast majority of cases "utterance + pause" does not capture the routine ways that recipients monitor talk in progress. What it does recurrently catch is a particular sort of problem posed for speakers' (p. 179). When these problematic silences occur, she also notes that they tend not to exceed about 1 s in length (Jefferson, 1989). Longer silences are avoided: "Whatever one might mean by "waiting long enough," waiting beyond 1 s is waiting too long" (Jefferson, 1986, p. 179).

There was, however, no indication that most of the silences in the Garrwa conversations were "problematic for the participants" (Jefferson, 1989, p. 170), which is in contrast to most of the longer silences reported in Jefferson's "standard maximum silence" paper. There was little evidence that these kinds of silences were particularly "meaningful." They appeared to be indicators of normal conversation.

Thus, in order to determine whether the talk is truly slow-paced, and unusually long gaps are regularly occurring, possibilities such as problems of various kinds, or orientation to another activity, need to be ruled out. In fact, some of the silences in the conversations studied for this project could be explained by interactional features of these kinds. Many others, however, could not. In the sequences of slow-paced talk we are describing, gaps are the norm, and there is little evidence of anything problematic or unusual in the talk. In fact, even where the slow-paced talk occurs around "problematic" activities, the talk is conducted even more lethargically than similar situations reported in the literature, with regular gaps far exceeding Jefferson's "one second" metric.

Rules, Gaps, and Lapses

The focal point for this paper is the nature of gaps of silence between turns, and the interesting case is what happens if current speaker does not continue, as this is the point at which gaps between turns emerge (cf. Wilson and Zimmerman, 1986). As Sacks et al. (1974) state, these turn-taking rules have a number of consequences for the conduct of ordinary conversation, including an orientation to the minimization of gaps between turns at talk. For example, if self-selection is used, then the incoming speaker is constrained by the possibility of current speaker continuing under 1c, as well as possible competition from other self-selectors, so an early start—as early as possible in the TRP—is necessary to assure speakership. As Moerman (1988) puts it, 'there is some pressure upon a person who wants to speak next to come in a little before … a possible end. Moreover, if he doesn't come in *now*, he may not get to come in *next* (and) an aspiring speaker who doesn't get to have his say *next*, might *never* get to have it' (p. 20). In a similar vein, Fox (2007) notes that 'speakers and recipients in real-time conversation have immense time pressures on them … recipients must be ready to start up a turn which is in some way responsive to the current turn, without delay, as

[3]While Kendrick and Torreira (2014) reported that dispreferred responses occur more commonly than preferred responses after a longer gap of silence, they also found that the most frequently occurring dispreferred and preferred responses occur after gaps of a similar length.

soon as the speaker has come to possible completion of current turn' (p. 314).

As Sacks et al. (1974) say, 'The components and the rule set, in organizing transfer exclusively around transition-relevance places, provide for the possibility of transitions with no gap and no overlap' (p. 708). Notwithstanding this provision, and Moerman's and Fox's observations above, silences can occur, and it is the very optionality of rules 1b and 1c—no speaker is obliged to self-select, nor is a current speaker obliged to continue if no other self-selects—that allows for the possibility of silences that can grow into extended gaps, and ultimately into lapses in the conversation, which is when participants disengage from each other. As Sacks, Schegloff and Jefferson put it, 'discontinuities occur when, at some transition-relevance place, a current speaker has stopped, no speaker starts (or continues), and the ensuing space of non-talk constitutes itself as more than a gap—not a gap, but a lapse' (p. 714). Note also that lapses can only occur after a current speaker has chosen not to continue when no other speaker has self-selected. No lapse can properly occur under the "current speaker selects next" provision[4]. As the authors point out, in such an event, 'a silence after a turn in which a next has been selected will be heard not as a lapse's possible beginning, nor as a gap, but as a pause before the selected next speaker's turn-beginning' (p. 715). In contrast, if no next speaker is selected, and no other speaker self-selects, and further, current speaker elects not to continue under 1c, then "a series of rounds of possible self-selection by others and self-selection by current to continue—rules 1b and 1c—may develop, in none of which are options to talk exercised, with the thereby constituted development of a lapse in the conversation" (p. 715)[5]. The implication of this is that in turn-by-turn talk participants are under considerable pressure to produce their turns early, within the transition space, but there are also provisions in the rules for turns to be delayed, which ultimately can account for lapses in a conversation.

Data and Methods

Our corpus consists of five conversations recorded in two remote Aboriginal communities in Australia's Northern Territory, near the Gulf of Carpentaria. Four conversations were recorded in the small town of Borroloola, which has a population of about 1000, the vast majority of whom are Indigenous Australians belonging to four different language groups. These conversations were audio recorded only[6]. The fifth conversation was recorded at Robinson River, a Garrwa (Aboriginal) community with a population of about 250 about 2 h drive south of Borroloola. There were five principal participants in these conversations: two elderly Garrwa women in Borroloola (Tina and Ellen[7]), and three elderly Garrwa women in Robinson River (Daphne, Hilda, and Katelin). The Borroloola data were mostly recorded on the veranda of a cabin, the second author and occasional passers-by entered the conversations on a few occasions. The Robinson River data features three elderly Garrwa women who were sitting on the ground on the front porch of the house of one of the women. We call this the "Porch" data. The recording, which lasts for over 2 h in total, was initially set up by the second author to record interactions between fluent Garrwa speakers and children who are not fluent in Garrwa. After about 20 min, the children leave (or are told to leave) by the elderly women, leaving the three of them alone on the porch. It is at this point that the task of recording Garrwa language ceases to be the focus of the talk and the topics turn to matters such as planning future hunting expeditions, complaining, reminiscing, and interacting with other residents as they pass by.

[4]Whilst lapses do not occur after "current speaker selects next" has been applied, it can be the case that gaps open up, and in these gaps it may be that, in the absence of a response by the selected speaker beyond the "normal" transition space (cf. Stivers and Rossano, 2010), another speaker may respond (Stivers and Robinson, 2006). In the cases these authors present, the gaps are mostly less than a second before a non-selected speaker responds. In some other cases, the selected speaker may be engaged in some other activity, contributing to the non-response. A central point the authors make is that there are two competing preferences at work when "current speaker selects next" is applied: one for the speaker so selected to speak, and a second for the action implemented by the current speaker's turn, e.g., a question, to be responded to, e.g., with an answer (if necessary by another speaker if the selected speaker cannot or does not answer). They say in relation to these two competing preferences that "the rule (i.e., 1a) is incrementally relaxed, suggesting an additional order of organization exists past the TRP which is laminated on top of the existing rule" (p. 391).

[5]We are not sure how one can determine that "rounds" of possible self-selection may develop (cf. Wilson and Zimmerman, 1986). An alternative possibility is that once the transition place has passed, and no next speaker has been selected under rule 1a, any speaker can self-select. This is, of course, a matter for empirical investigation, and is beyond the scope of this paper. However, cases such as the following (taken from the first author's data) raise interesting questions.

```
1   Liz:     =Oh;= that's good.
2            (6.5)
3   Mel:     °°(whoohh hehh)°°
4            (8.6)
5   Mel:  ->  [Tom-] (.) [Tommy-]
6   Liz:  ->  [D'ju] [D'ju wa-]
7   Liz:      Sorry-. Mel d'yu wan' that o nion.=
              bicoz-
8             (0.7) °I- (.) > don' thing I'll< (.) >be
9             able ta< digest it.
```

Liz and Mel are at the dinner table. After line 1, a silence ensues, grows, and becomes a lapse, though there is a sigh after six-and-a-half seconds. After no talk for about 16 s, both start to speak simultaneously. Both drop out after one beat of talk, there's a micro-pause, both begin to talk again with identical starts, this time they go on for two beats, and again both drop out. Liz then gets the floor at the third attempt, starting with an apology. The question is what prompted such precision timing in their starts after such a long silence. Preliminary findings by the first author suggest that some body coordination occurs before such post-lapse simultaneous starts. For the arguments of direct relevance to the current paper, such instances would suggest that after a certain length of silence, any speaker may start at any time. Whether that is the case immediately after closure of the transition space, or after a second or so (cf. Jefferson, 1989), or after some other metric remains to be determined. If this turns out to be the case, it would mean that the priority given to self-selection (rule 1b) over current speaker continues (rule 1c) would at some point be superseded.

[6]We acknowledge the shortcomings of working solely with audio recorded data. The Borroloola recordings were originally made for linguistic analysis, and only later used for conversation analysis. For this reason, we use them only when there are no good illustrations of the points we are making from the video data. Given the relative dearth of Conversation Analytic studies on endangered, non-Western and non-industrialized language groups, we feel justified in including some audio-only data.

[7]To preserve anonymity, names of participants have been changed.

In all of the conversations much of the talk a mix of Garrwa, a local variety of Kriol (a local creole language) and Aboriginal English, reflecting normal patterns of community multilingualism. A total of about 35 min of the Borroloola conversation, and about 26 min of the Porch conversation have been closely transcribed. In the collection of extracts from these conversations we had over 400 speaker transitions, and within this collection we analyzed more closely about 160 of the longer gaps. Four of the five women acted as informants and teachers of the language as part of the transcription process.

Where we have video evidence, most (but not all) of the long silences do not occur when participants are engaged in non-talk activities, such as those Jefferson (1989) points to explain the extended silences in her data (including examining a train timetable, scanning the surroundings in a neighborhood block party, or writing down an address). On some occasions in our data, something is going on in the environment that might hold the participants' attention, but on others the speakers are sitting around doing nothing apart from talking.

We use Conversation Analysis to analyse the silences within sequences of talk (Sidnell, 2011; Sidnell and Stivers, 2012). These elderly women were among those who taught the second author the grammar of Garrwa, and features of morphology are included in the transcriptions to show turn construction features. The women also provided us with some ethnographic and local contextual information where we needed it to understand what was happening in the interactions, for example about the collection of wild honey ("sugarbag").

We measured the pauses by manually locating them through the inspection of waveforms in Audacity (Audacity Team, 2013), and these figures were then rounded up or down to the nearest tenth of a second. The response offset is measured between the last element of the first turn, and the first element of the responding turn, whether it be a particle such as *well* or *uhm*, or a lexical item. In breaths or clicks were also included as part of the second turn.

Results

In the data we have examined for this paper, we commonly found stretches of talk such as extract 1.

(1) Porch:2.1:866:PD2.

```
866  Kat:        Jurarrba ngayu ngawukuku.hh
                 Hot   1Sg   pregnant.belly
                 I'm angry
867          ->  (1.3)
868  Dap:        Jurarrba >ninji< ngawukuka.
                 Hot      2Sg     pregnant.belly
                 You're angry
869          ->  (2.3)
                 ((Daphne drinks from bottle))
870  Kat:        'ana:nkuny' wawarrany'.
                 DEM-DAT  child-DAT
                 With those kids
871          ->  (2.2)
872  Dap:        Barri balba yali; = bukamba na, (0.3)
                 Barri go  3Pl-PAST  all  NA
                 They've all gone
873  Dap:        wi:jba 'li  k'ngkarr'; = s:choolyurri.
                 return 3Pl-PAST up  school-ALL
                 they've gone back up to school
874          ->  (4.5)
875  Hil:        Mm: ^hm.
876          ->  (0.5)
877  Dap:        Barri[wa.
                 Barriwa
                 Finished/Anyway
```

Extract 1 occurred at Robinson River while the three women were sitting quietly alone, mostly without eye contact, though occasionally turning to each other. Two of the women are sitting next to each other, facing the house at an angle. The third is sitting at an angle of about 90° to the other two, behind them and facing away from the building, and two (Katelin and Daphne) remained in these positions for the whole of the 2 h of the recording, with Hilda arriving after about 40 min. To obtain eye contact with each other, they needed to turn their heads 30° or more. The video for extract 1 shows Katelin fiddling with small unidentified objects in her lap, Daphne with a bottle of drink, and Hilda stroking a coolamon[8]. At times such activities lead to what Goodwin (1981, p. 106) has called "activity-occupied withdrawal" from the talk, by which he meant activities such as writing, preparing food, grooming, attending to equipment or other artifacts (Goodwin, 1981, 1994, 1995a). In the Borroloola and Porch conversations such activities mostly did not appear to disrupt the talk, occurring during both talk and silences. Some other activities, however, namely drinking, smoking, coughing and nose blowing, were potential talk inhibitors[9].

Between each turn in this extract (except the last two), there are gaps of silence of between 1.3 and 4.5 s, and there is no evidence, either auditory or visual, that the participants experience any problem with the talk, nor are they engaged in other activities, apart from fiddling with small objects in their hands and, in line 869, during the 2.3 s pause in line 869, Daphne drinking from a bottle.

Orienting to the Rules of Turn-Taking

So what is happening here to the notion of orientation to TRPs and gap minimization? An initial point is that this has nothing to do with a different set of turn-taking rules. As Sidnell (2001) found for Caribbean Creole conversations, the turn-taking system in the conversations of these Garrwa speakers has the same fundamental organization as that described for American and British conversation. These women routinely speak when selected. They self-select when they are not selected. They sometimes respond with precision timing at TRPs, orienting to possible completion points of TCUs. In extract 2, Tina selects Ellen as next speaker by asking her a question (rule 1a). Ellen responds in "unmarked next position" (i.e., with one beat of silence) when she is asked a question by Tina.

[8] A coolamon is a vessel made of bark or wood for carrying water, babies etc.

[9] Other activities in the Robinson River conversation which either were seen to accompany talk, or in our judgment could do so, were: shaking an asthma inhaler, waving away flies, wiping away sweat with a towel, rolling small objects on the lap or thigh, putting objects into a handbag, scratching, brushing hair, and opening or closing a bottle, as well as peering out into the surroundings.

(2) Garrwa3-.20.8.03:V3:37:0′30″

```
37  Tin:    YE:UH? (.) bak' wi:nj'wa barri nyulu-.
                       and  where    BARRI 3sg
            Yeah, and where (is) he?
38  Ell:    'E  -la: pub
            He   LOC pub
            He's at the pubF
```

In extract 3, in line 60, Tina selects Ellen as next speaker under rule 1a, by initiating repair with "Who." Ellen latches her response to this repair initiation. It can also be noted, however, that the gap between Ellen's question and Tina's repair initiation is almost twice as long as the average for English conversations (Kendrick, 2015).

(3) Garrwa3-10.10.03:V3:58:1′00″

```
58  Ell:    >Wanya'kurr' 'al';<balb'< nayi barri:
            .hh-u
             What-ALL    3pl   go     here BARRI
             Where are they going here
59          (1.4)
60  Tin: -> Who:.=
61  Ell: -> =For S:nora (boy) dat  bo:y blang:h
             For Sonora, for that boy there
```

Self-selection under rule 1b can also occur with no gap. In extract 4, Tina announces that she had been dreaming, which does not select Ellen as a next speaker. Ellen then self-selects, on a non-related topic, latching her own remembering to Tina's turn.

(4) Garrwa:8.9.03-V3:18:0′25″

```
22  Tin:    ↑Dreamin' ngayi.=
             Dream    1sg-PAST
             I was dreaming
23  Ell:    =Ye::ah. (.) -an' dem- (0.6) >↑(Y)ee:
             Yeah         and those          (yeah)
24          wi'baki'      (wi'bayi)    hh
            fall.down-IMP fall.down-PA
            Yeah, and those- (yeah) they should have
            and did fall down
```

There are also cases of turn starts in terminal overlap, that is, early in the transition space. In extract 5, Daphne observes that two people are cooking a kangaroo, and Katelin notes how they are cooking it, overlapping with the last part of Daphne's turn.

(5) Porch2.8:614:0′30

```
615 Dap:    Kularra bul' li[rrkidi.
            South   3Du  young.kangaroo
            Two people (cooking) the young kangaroo
            to the south
616 Kat:                   [E boi:lim >^billy?
                            They're boiling it in a
                            billycan
```

Another example of self-selection under rule 1b occurs in extract 6. In this example, in which they are talking to a man who is passing by, both Tina and Ellen self-select in 272 and 273, after 0.2 s, which is the later end of the transition space. Tina is first starter, which, under rule 1b, means she has rights to the floor. Ellen, the second starter, drops out, but then restarts (line 274) immediately Tina finishes her turn, latching her talk to the end of Tina's talk, just as predicted by the Sacks et al. (1974) turn-taking rules.

(6) Garrwa3:20.8.03:V3:268:5′05″

```
268 Ell:    B:arri
            This morning
269         (0.6)
270 S:      Y:eh.
271         (0.2)
272 Tin: -> 'E M[US:' BE UP DE:RE.=
            He must be up there
273 Ell:       ['E-
274 Ell: -> ='E migh' be up dere  la' Shirley.
            He might be up there at Shirley's camp
```

A further example of self-selection at a point of possible TCU completion occurs in extract 7, which is from the Porch data. Daphne is asking a passing girl to get her mother to bring some fish and chips. Daphne gets two responses to her request, one from the girl, and subsequently one from one of the other older women, Hilda.

(7) -Porch-2.10:1220:1′00″

```
1220 Dap:    You go- go ba:ck;= and yuh tell
1221         yer mum:;= I wanda fish an' chi:ps.
1222         (0.7)
1223 Girl:   She ˇgawn ^ta:hn.
             She  gone to town
1224         (0.4)
1225 Hil:  -> E  bin ^ga:wn;= [ˇtuh do]cter.
             She's gone, to the doctor's
1226 Dap:  ->                [ When. ]
1227         (0.3)
1228 Girl:   Toda:y:.
```

At a point when the second of these responses, by Hilda, is possibly complete, after "e bin gawn" (she's gone), Daphne asks a follow-up question, "when."[10] However, it turns out that there is more to come in Hilda's turn: "tuh docter" (to the doctor). This is an example of what Jefferson (1984) calls latched overlap, which occurs because Daphne has not predicted the extension of Hilda's turn.

In these conversations we also find examples of the occurrence rule 1c, in which a current speaker continues when no other speaker has self-selected under rule 1b. In extract 8, Katelin requests that the other two women start talking more as she has been doing most of the talking so far (this is for the benefit of the recording of the talk), as she is tired. There is no evidence in the video that she is directing her request at only one of the women through, for example, gaze selection. Katelin comes to a TRP at the end of line 846, with grammatical, intonational and pragmatic completion (cf. Ford and Thompson, 1996). There is no response within 0.4 s, at which point Katelin continues her turn with an account for her tiredness.

(8) Porch:2.8:845:3′50″

```
845 Kat:    *Ya:nyba^kiji*;= ninji marda,= ngurujba
             nga:yu,=
             Talk-IMP        2Sg   also    tired  1Sg
846          ^du    yany:bikanyi?^
                    yanybikanyi
             (too)  talk-KANYI
```

[10]It is possible that Daphne is responding late to the girl's turn in line 1223, but even were this so, the precision timing of Daphne's "when," latched to a point of possible completion, remains.

```
                You talk now because I'm too worn out to
                talk
847         -> (0.4)
848  Kat:      Yinybayi  nga';= ^all: day wula[:ni.
               Talk-PAST 1Sg              day.before
```

In extract 9, from one of the Borroloola conversations, we present two examples of rule 1c, in this case both clashing with 1b. This clash occurs because a self-selecting speaker comes in rather late—at the end or even beyond the "normal" transition space—at the same time as a current speaker elects to continue.

(9) Garrwa3:20.8.03:94:1′40″

```
94   Tin:    Wan:yinyini barr'?
             What-ERG    BARRI
             Who (did the chasing)?
95           (0.3)
96   Ell:    Nana barri kaburri. hh
             That BARRI young.man-ERG
             That young man
97           (0.6)
98   Tin:    Yindi¿
             Really?
99           (1.0)
100  Ell:    ·HHH E: muddehr HHHH=
                  His mother
101          (0.2)
102  Ell: -> [Mudder; hhh]hh
103  Tin: -> [Im -mudder.]
              His mother
104          (1.6)
105  Ell: -> [Im yu:-]
106  Tin: -> [E  m:ud]der deh.
              His mother there
107          (0.2)
108  Ell:    Im: u khasim dat  ngi:la na¿
             He    chase  that girl   NA
             He is chasing that girl
```

In line 98, Tina produces a newsmarker, which appears to be seeking confirmation that it was the young man who did the chasing. In response, Ellen makes reference to his mother[11]. After 0.2 s, which would be at the end of the normal transition space for British and American conversation (Wells and Macfarlane, 1998), there is a simultaneous start with Ellen repeating "mudder," and Tina producing a confirming repetition, "im mudder." As these two turns were produced simultaneously, it is now equivocal who, at the simultaneous start that follows in 102-3, is current speaker and who is next: in effect both are current speakers, so both, after a long transition space of 1.6 s, elect to continue, Tina continuing the same sequence with "E mudder deh," and Ellen moving on to something new. However, Ellen drops out, Tina completes her turn, and 0.2 s after Tina finishes, Ellen restarts the turn she had abandoned.

In the final example in this section, extract 10 presents a sequence that can clearly be seen as potentially problematic, in the sense that a question is asked that never gets answered. However, there is no evidence from the talk or from the video that the participants orient to it as particularly problematic, not even Hilda, who asks the question.

(10) Porch2.7:545:4′30″

```
545  Hil:    Wanja' 'lu;= Black[fella Spring;= wanya,
              Who   3Pl                        who
              Who's that at Blackfella Spring, who?
     Dap:                    [Glances at Hilda
546         -> (7.7)
547  Hil:    Wanya nan:'a yi;= [karrurru ju:nu,
              Wanya nanda ( )   karrurri junu
              What  that        east     don't.know
              What's that (place) in the east, wherever.
     Dap:                       [Turns head towards
                                 Hilda
548         -> [(2.6)
     Dap:    [Turns head away from Hilda
549  Hil:    Nuyiburri˘°nanyi°.
              Nujburrinanyi
              Valley-ABL
              From the valley
550          (1.6)
551  Dap:    Ngarrangarra ngana;= ngadara[nyi   barri,]
              Ngarrangarra ngana   ngadarawanyi  barri
              Burn         1SgACC  sun-ERG
              The sun is burning me
```

Hilda twice pursues an answer, following the rule 1c according to which if no other speaker self-selects then current speaker may continue. This happens in lines 547 and 549, but after these two attempts, she gives up. Throughout this extract, Hilda is gazing to her right at 40° toward Daphne, while the latter is fiddling with the cap on her bottle of soft drink, and the sound of gas escaping is audible. This suggests she may be preoccupied, (as is Katelin—not the addressee—who is brushing her hair). However, fiddling with a bottle cap is not an activity that would necessarily inhibit Daphne from answering Hilda's question. After the first question, Daphne glances at Hilda, but there is no response during a gap of 7 s. There is also evidence that Daphne has heard the second question, because toward the end of line 547, she turns her head toward Hilda, holds her gaze briefly, and then returns it to her drink bottle during the 2.6 s silence that follows. Then there is increment to this question in line 549, "Nuyiburri nanyi" (*from the valley*) and again nothing for 1.1 s, which is when Daphne does produce a turn, which, however, is not a response to the question, but a complaint about sitting in the sun. It is also notable that Hilda does nothing to show that she finds the lack of an answer to her question problematic, apart from twice pursuing the answer before dropping it.

These examples demonstrate that these Garrwa women can, and regularly do (though as extract 10 shows, not invariably), orient to projectably complete units of talk, the turn-taking rules, and transition places, as explicated in Sacks et al. (1974). What, then, is going on when there are regular long silences between turns in these conversations?

[11]It is unclear to us who the "mother" is who is being referred to here, and why she is being referred to. It is a common practice in Aboriginal communities for reference to persons to be indirect or vague (Blythe, 2009; Garde, 2003). This phenomenon can be explained, at least in part, because they live in small communities in which everyone knows everyone else, but also because there are kinship related taboos on naming certain persons directly. Be that as it may, following Tina's repair initiating "Yindi" in line 98, there is a sequence in which the reference appears to be resolved to the satisfaction of Tina. The point we are making in this extract about turn-taking practices following Sacks et al. (1974) rules holds despite the reference to "mother" remaining unclear.

Accounting for Long Silences and Expanded Transition Spaces

We have reported so far that in our data longer gaps between turns are common. There are, however, striking differences in the length of silences between the "current-speaker-selects-next" (1a) and "next-speaker-self-selects" (1b) techniques. This may not be surprising, but this is another feature that shows that Garrwa speakers are conducting their conversation in a similar way to English conversationalists. When a current speaker selects a next, the response tends to come relatively quickly, though with a longer delay than has been observed in other languages. In all of the examples examined from our Garrwa corpus, silences after speaker selection occur regularly up to 1.5 s (Mushin and Gardner, 2009). In contrast, when no next speaker has been selected, the gaps can be much longer. In this section, we examine some extracts in which current speaker selects next, and following that, we consider some cases in which no next speaker has been selected.

Silences After Selection of Next Speaker

As was noted above, silences occurring when a current speaker selects the next are of a very different order to silences when the next speaker self-selects. According to Sacks et al. (1974), if a current speaker has selected the next speaker, there is an obligation for the selected participant to speak as early as possible at the next transition space. If no next speaker has been selected, then no such obligation exists.

On some occasions the TRP may be "expanded," for example because of activity-occupied withdrawal, or a dispreferred response. However, on many occasions in the Garrwa conversations, when a current speaker has selected a next under rule 1a, there is a gap preceding even a preferred response. Regularly, but not always, these silences are "filled" with relevant gestures, such as head nods which precede the talk, and there is thus no delay in the response in such cases. On others, there is no talk-supporting activity. There is a delay of 1.4 s in extract 11 between the question and answer.

(11) Porch:2.1:1002:IR-4:1′35″

```
1004 Dap:   >An' wanyi kuyu  nan'  yiliburr'-.<
             And what  bring that  waterlily
            And who brought that waterlily?
1005      -> (1.4)
1006 Kat:   ^Ya:^lu,= minjil'  yal:';= jila karrina
             3Pl      CONJ-HAB 3Pl     go   east-ABL
1007        Win:mirrinanyi,hh
            PLACE.NAME-ABL
            They used to, they went from the east,
            from Calvert Hills Station
```

This is a case in which there is no gestural support of the talk. Katelin and Hilda are both looking at Daphne whilst she is asking the question. During the question and the 1.4 s silence, Katelin and Daphne have eye contact, though Katelin is fiddling with her collar, which is not an activity that necessarily precludes simultaneous talk. There is nothing in this sequence that indicates any trouble, nor does Katelin appear to have her attention on any other matter. The answer, when it comes, is preferred, though it is expanded (and it is expanded further beyond this short sequence)[12]. This contrasts with a typical preferred response in "Western" talk, where such a response tends to come quickly and briefly, without accounts or other expansions (Schegloff, 2007: 67ff).

The broader context of this adjacency pair is that it occurs during a reminiscing sequence about how in the old days they used to collect waterlilies to eat. This may help explain another feature of this sequence, namely that this is on the face of it an information-seeking question, which in English mostly attracts a brief, phrasal response (Fox and Thompson, 2010), but here we have an extended response—which is further extended beyond this adjacency pair (not shown), so this WH-question could be seen as a prompt for extending the reminiscing. Planning for an extended response may be a factor in the delay of 1.4 s.

Extract 12 is from the Borroloola corpus. There is no video to support the analysis, but this is included as an example of another delayed answer with a preferred response, with no perceptible trouble.

(12) Garrwa-9.10.03:V3:99:1′30″

```
102 Ell:  ->  An' where I bin:.
              And where was I?
103       ->  (1.0)
104 Tin:  ->  Yuh ovuh ↑*deh*.
              You were over there.
```

In extract 13, the silence cannot be accounted for even in part by non-talk activities or delays associated with a dispreferred response. Hilda repeats an answer that she had already provided once, namely that wild honey can be found at Hubblestrap. Daphne and Hilda have eye contact, with Daphne's neck "torqued" almost 90° toward Hilda (Schegloff, 1998). They are clearly focused on talking to each other.

(13) Porch2.10:1282:2′00″:IR-5:2′13″

```
1279 Dap:   AY: MUM.=
            Hey mum
1280 Kid:   ='e:uh.
1281        (0.7)
1282 Hil:   °Jungku° [wanyi barri;= lang' Hu:bblestrap=
             Sit      what          LOC   PLACE.NAME
1283 Dap:            [wa-
1284 Hil:   =°nangangi°.
              nangangi
              3SgDAT
            It's at Hubblestrap
1285        (0.3)
1286 Dap:   Yang:kawa-;
            How-WA
            Whereabouts
1287      -> |-----------------(1.7)-----------------|
      Dap:  |Looks up, turns head slowly to the left|
1288 Hil:   | Righd where dem ^gr:id | dere ba^^yu[ngu?=
                                        west
      Dap:  | Single nod down then up|
            Right where that cattle grid is to the
            west
```

[12]The vagueness of "they used to" in response to a "who" question in line 1006 reflects a common practice in Indigenous Australian talk of vague person reference (eg., Garde, 2003; Blythe, 2013). This may in part have to do with the small, close-knit communities in which everyone knows everyone else. It may also have to do with taboos, where naming of certain kin is forbidden (Blythe, 2013). This issue is beyond the scope of this paper.

Between the answer that Hilda had given in 1225 in extract seven and this sequence, the three women had been talking to two boys who had arrived where they were sitting. Daphne then turns her head sharply toward Hilda in 1279 with a summons (the "Mum" in 1279 is a term of address directed at Hilda), which has the effect of Hilda repeating her answer, that the sugarbag is at Hubblestrap, and Daphne then asks for more specific information with "yangkawa" (whereabouts). There is a delay of 1.7 s between the repair initiation in 1286 and the response in 1288. What happens in this silence is fully oriented to the answer. Hilda turns her head slowly in a westerly direction (away from Daphne), and this takes up the whole 1.7 s, and then nods in that direction as she says "Righd where dem grid." There is no hurry to start talking. The head turn prepares for the answer, and is accomplished prior to the verbal response. This languid response contrasts with the general practice in English (Goodwin and Goodwin, 1987) and Japanese (Kita and Ide, 2007) conversation, where it is most usual for the gesture and talk to occur very quickly after the prior turn, with the gesture slightly foreshadowing the words (Streek, 1993).

In the next extract Daphne urges the other two to hurry up so they can leave. Her directive in line 894 receives no immediate compliance from the other two. Indeed, first Hilda and then Katelin concur, but with substantial gaps of 1.2 and 2.4 s, respectively.

(14) Porch:2.8:894:4′40″:IR-3:4′12″

```
894  Dap:   Wajakiy'  (janungkala );= ˇbalba *mili*ˇ.
            Quick-IMP (FUT-1DuIncl}   go    INTENS
            Hurry up, (we two) have got to go
895       -> (1.2)
896  Hil:   ^We ˇgodda go;= ^jarr:ba¿
                            eat
            We've gotta go eat
897       -> (2.4)
898  Kat:   ^Wa:dambiji;= ^mukunˇja:nanyi¿
             Feed-PURP     dinner-DAT
            To eat dinner
899         (1.0)
```

This is an agreeing sequence, that is, each turn is a preferred next: both recipients say they want to go to eat something, but neither shows any sign of complying with Daphne's directive by getting ready to leave. There is no observable ambient or interactional reason visible in the video for the delays between these turns. However, rather than the immediate or even early responses to preferreds that are reported in the literature (Levinson, 1983; Pomerantz, 1984a; Schegloff, 2007), there are substantial delays here. Daphne is waving away flies and then she picks up a small object and shakes it: but she does this *after* she has spoken. Hilda is stroking a coolamon throughout this sequence, but this is not an activity that interferes with the ability to talk. Daphne is urging them to quick action, so what better illustration of the slower pace of the talk could there be than this languorous hurrying up?

We can see from the examples presented in this section that sometimes when the current speaker selects next, the response occurs relatively promptly, mostly within about a second-and-a-half (see Mushin and Gardner, 2009 for a more detailed discussion), whilst others (not presented) have delays of less than a second. The metric with a maximum of about a second-and-a-half is similar to what Scollon and Scollon (1981, p. 25) claimed for Athabaskan. This metric is about half a second longer than Jefferson (1989) found for the American, British, and Dutch conversations she studied, where there was a "standard maximum silence" of about 1 s (0.9–1.2) for various kinds of silence. The longer silences she found could be accounted for by activities the participants were engaged in that interfered with the flow of the conversations. For these Garrwa conversations, there may be grounds for amending Jefferson's observation about waiting for 1 s to: "Whatever one might meant by "waiting long enough," waiting beyond *one-and-a-half* seconds *after one has been selected* is waiting too long" [adapted from Jefferson (1986), p. 179]. In the next section, we shall discuss gaps of silence between turns where there is no selection of next speaker by the current speaker.

Silences Before Self-Selection by Next Speaker

The metric of a maximum silence of one and a half seconds appears to apply only to turn transitions in which next speaker has been selected by current speaker, but not to self-selection in turn-by-turn talk. This can be partly explained by the fact that where there has been no selection of next speaker, there is no obligation for anyone to speak. A gap may ensue, and can extend until there is a lapse in the conversation and speakers disengage. In many cases in the Garrwa conversations, however, inter-turn silences of several seconds occur without any apparent orientation to a problem in the talk, nor any indication from changes in body posture that the conversation has lapsed. Some of these gaps can be explained in the same way as those discussed earlier: sometimes non-talk activities, or dispreferreds, or the ends of sequences and topic attrition occur in conjunction with longer silences. On other occasions, however, such factors do not appear to elucidate the silences.

In extract 15, which is from the beginning of the Porch conversation, there is very little non-talk activity from the three women: Katelin scratches her foot, and Daphne appears to wave away a fly, but otherwise they are sitting and looking mostly straight ahead, without eye contact. They do not appear to pay much attention to the barking dog, except for Hilda's question about it in line 8.

(15) Porch:2.6:001

```
            ((Dog barking continuously in the
            background))
1    Kat:   Yarrijba yalu walukuku wanananba.
            Put      3Pl  humpy    all.around
            They built the humpy[13]  all around
2         -> (5.2)
3    Kat:   >Jungkuyi mili   nani baluwayi
             Sit-PAST INTENS that look.attentively-
             PAST
4           junu.
            don't know
            He was sitting down looking attentively,
            whatever
5         -> (4.3)
```

[13] A humpy is a small temporary shelter.

```
6     Dap:   Baya
               Bayinya
               (Round the) back
7          -> (1.6)
8     Hil:   >Wanya nani wandu(wab)a.<the e
               What  that (bark)
               What's that (barking)?
           -> (1.7)
9     Kat:   °Badadajb' yalu kingkarrinya,°
               Badadajba yalu kingkarrinya
               Come(pl)  3pl  up-ABL
               They're coming up
10         -> (0.5)
11    Hil:   Wakanya
               Pig
12         -> (1.0)
```

In the above extract, there is no topic that they are pursuing. They've been sitting around for a long time before this point in the conversation. There is no strong engagement. They move in desultory fashion from topic to topic, commenting on what is going on around them—a dog barking, a group of people approaching. Reference by Katelin to a "humpy" (temporary shelter) is followed by 5.2 s of silence. The talk in lines 3–4 is followed by a silence of 4.3 s. Daphne then self-selects with an observation that something is round the back. 1.6 s later, Hilda asks a question which does not get answered. 1.7 s after this, Katelin then observes and comments on some people approaching. Half a second later Hilda mentions the presence of a pig nearby.

In contrast to extract 15, where there is little topical continuity between the turns, in the next extract, the gaps—even the 7.6 s in line 115—are all between turns that are coherent self-selecting contributions in the flow of the talk.

(16) Garrwa2-9.10.03-V2:111: 1′45″

```
111 Tin:   >All duh-< (0.4) whatsihname bin; (0.5)
             fiximup
112        ↓in*side im.*↓
           All the - whatsitsname - fixed him up inside
113     -> (1.2)
114 Tin:   <Whunim  blee:din'.> (.) >kurda-¿<
           When he was bleeding, poor fellow
115     -> (7.6)
116 Tin:   ↑Ny:ulu m'li  karij-kurr~i:~.
            3sg    again (want.to.do)-DS
           He (wanted to do) it again
117     -> (1.5)
118 Tin:   E jealous (it's his) ba:ba:.
           She's jealous (of her) brother
119     -> (1.7)
120 Tin:   ↓for dem  ↑ku:ku.
            for those grandchildren
121     -> (1.1)
122 Ell:   ↑Dey ↓u-wan' jung'u.=↑talk la im;=
123        drunkaman y'knoo¿
           They want to sit and talk there to the
           drunk one, you know
124     -> (1.0)
125 Tin:   Uh h[uh¿
126 Ell:       [E bin ledim (avim); ba:rdibardi¿
                She let them (    ), old woman
127        (1.2)                                  —
128 ?:  -> ((sniff))                          (4.8)
129        (3.1)                                  —
130 Tin:   [(   )]
131 Ell:   [Yabu ] f`r -e) ↓nevuh] ↑come↓for ↑da:nce,
           That one never comes to the dance
132     -> (2.0)
```

In the first part of this extract, all of the talk is by Tina, with no hearable responses from Ellen, although there appears to be no reason why Ellen could not have self-selected. This is not a storytelling, although Tina's first turns refer to a recent event in the town. They have been exchanging views about people. None of these turns selects a next speaker, there is no competition for the floor, no sense of having to get the next turn in "now and not later," and neither participant shows any urgency in producing a next turn. Two of the silences could be analyzed as intraturn silences, namely the ones in line in line 113 and 119, each of which is an increment to Tina's prior talk. But even so, the prior talk in each case ends at a possible TRP, and thus speaker change is relevant. For the talk that comes after the 7.2 s silence in line 115, however, it is harder to analyse this as an intraturn silence, as what precedes it is potentially complete, and what follows is not an increment. The talk flows topically, and could have been produced as a coherent multi-unit turn without any silences. Furthermore, there is no aural evidence of any other activity during this silence.

In this section we have looked at inter-turn gaps that occur when no next speaker has been selected. We have found that there are numerous gaps in these positions, and some of them are very long. The extracts presented here are ordinary conversation, and some of the languor of the talk here can be explained by non-talk activity or dispreferred actions. However, most of these non-talk activities are grooming or fiddling with objects, activities that could accompany talk, in contrast for example, to reading or writing, which require more focused attention. In very many sequences in these conversations, the inescapable observation is that this talk is inherently languorous, and there is no attempt at minimization of gaps in the way that has been described for English conversation, for example in Sacks et al. (1974). Very long silences, or lapses, are not frequent, as we have only three silences of longer than 10 s in our corpus[14]. Furthermore, gaps of more than 2 or 3 s are not common. We do find overwhelming evidence of an orientation by these speakers to the rules of turn-taking as presented in Sacks, Schegloff, and Jefferson. The difference from most published literature on turn-taking is one of the "value of variables" (Schegloff, 2007, p. 74). It is a difference to one of Sacks, Schegloff, and Jefferson's gross observations: "transitions (from one turn to the next) with no gap and no overlap are common. Together with transitions characterized by slight gap or slight overlap, they make up the vast majority of transitions" (pp. 700–701). In many sections of these conversations, the majority of transitions are, in contrast, characterized by substantial gaps.

Extended Gaps in Conversation in Other Contexts

An easy explanation of the phenomenon we are reporting in this paper would be one of cultural difference: that Aboriginal Australians have a different "conversational style." Whilst we have found regular lack of gap minimization in these conversations, there is no evidence in the data of a different set of turn-taking

[14]Unfortunately these are all in the Borroloola audio-only data, so they cannot be checked for what the participants were doing during the silences.

rules for these speakers. As Sidnell (2001) noted in his study on Caribbean Creole conversations, "[t]here is, at time of writing, no empirical study which provides evidence that humans do conversation in a fundamentally different way" (1286). We provide no evidence for this either. Our study provides another example of a non-Western culture in which the fundamental organization of conversation appears to be "grounded in a species-specific adaptation to the contingencies of human social intercourse" (Sidnell, 2001, p. 1263). The difference is only that regularly there is a marked expansion of what counts, at least in Western conversation, as a normal transition place.

In fact, the slow pace of these conversations is not, on the evidence we have, culture–specific. We have some evidence that expanded transition spaces occur in "western" talk. In the data used by the first author for his work on response tokens (Gardner, 2001), there are examples of slow-paced conversations amongst Anglo-Australian couples who recorded themselves when they were at home alone. These are intimates, engaged in talk at times of day when the pressure is off, such as after the evening meal. There are times in these conversations that look and sound very similar to the Garrwa conversations. In extract 18, there are some very long gaps, including one of 4.3 s in line 163 following a question from Liz, that is, after she has selected Mel as next speaker.

(18) L&MC2ai-Languorous:144

```
144          (24.7)
145          ((dishes being cleared))
146 Mel:     °°(On) channel two°°; (0.4) °channel two:°.
147          (0.4)
148 Liz      °Mm°.
149       -> (7.7)
150 Mel:     °Oah°.(0.4) goe:s ta about te:n fiftee:n¿
151       -> (3.3)
152 Mel:     Then the ne:ws:,= so I guess (0.4) we
153          mighd even watch that.
154       -> (1.4)
155 Liz:     -nhhh
156          (0.7)
157 Mel:     Haven't seen the ne:ws fer a whi:le¿
158       -> (5.0)
159 Mel:     °We:ll;= I said I wozzen gonna wa:tch
160          the news any mo:re°,
161       -> ()
162 Liz:     >Why not-<.
163       -> (4.3)
164          ((newspaper rustling))
165 Mel:     >C'z ah j'ss< ra:ge about (the-) (0.5)
166          whole lotev (0.5) -cr:ap.
167          (0.9)
168 Liz:     °Mm:°.
169 Mel:     En: (0.2) negadive stuff; (0.5) no I
170          don'- need that, (0.5) (mode)¿
171       -> (3.9)
172 Liz:     >D'y'ink it's more negedive th'n
173          list'ning ta the ne:ws.
174       -> (1.6)
175 Mel:     Ye:hs.
```

In the early part of this extract, Mel appears to be reading the television schedule for the evening in a newspaper, an example of "activity-occupied withdrawal," with interspersed comments on programs. In the latter part, though, there are questions and answers, and the slow pace continues. The rustling newspaper in 164 might suggest that Mel is still engaged in reading—which would be an explanation for the long gap—but still, overall, the pace here is slow, and there is little attempt to minimize gaps.

In extract 19, Ike and Jan are driving, and are discussing who is going to have the car later.

(19) I&JW4a

```
39  Ike:    So I'll go back in la:der t'ni:ght;= °en do
40          couple uh hour's wo:rk.°=
41  Jan:    =↑↑Wh:a::::[t↑↑ ?
42  Ike:               [°an' drop thuh stuff back°.
43          (0.9)
44  Ike:    >↑WELL UH GOTTA DROP IT BA:CK<.
45          (1.3)
46  Jan:    >°Do't in th'° ↑mo:rning↑.<
47          (1.0)
48  Ike:    °W'll I haven' got a ↑ca::r.= [fuh two weeks°.]
49  Jan:                                  [°You  c'n  take]
50          th' ↑ca:r°.
51       -> (4.1)
52  Ike:    °You ↑need ↓thuh ca:r°.
53       -> (4.5)
54  Jan:    ↑↑'oo k' ↓go home by ↑↑trai:n?
55          ((yawning))
56       -> (3.5)
57  Ike:    So we go in t'-gether;
58       -> (2.3)
59  Ike:    >d'yuh mean<¿
60          (0.7)
61  Ike:    an' I:'ll take thuh ca:r;= (tay) drop you off,= en
62          (0.9)
63  Jan:    OR ELSE I C'D DROP YOU: OFF,=en::' (.) °go in:¿°
64          (0.6)
65  Jan:    *like* with:: *thuh other* ti:me¿
66          (3.0)
```

This conversation was recorded during a quite lengthy car journey. It begins with a proposal by Ike that he go into work in the evening, which gets a less than enthusiastic response from Jan. What follows is a series of proposals and counter-proposals and accompanying accounts. What is notable and relevant to this paper are the silences between turns in this sequence of up to 4 s.

The extended interturn silences of these Anglo-Australian conversations provide evidence that slow-paced talk is not restricted to cultures or societies such as the Garrwa people, or the Native American Warm Springs indigenous people that Philips (1983) reported on. We certainly have not found in these materials, as Philips (1976) claimed for the Warm Spring Indians, that turn-taking by their system was self-directed, or that anyone who wanted to speak did so and for as long as they wanted (Philips did not provide closely transcribed materials to back up her claim.). There are further situations in which talk characterized by expanded transition spaces is normal: second language classroom talk (from the first author's data), second language conversations (Wong, 2000), and perhaps most extremely in hypnosis sessions (Demosthenous, 2008). In the last of these, gaps of more than 10 s regularly occur between a hypnotist's question and a client's answer whilst in deep hypnosis.

What may be happening in these Garrwa conversations (and in the other interactions for which expanded transition spaces have been reported) is that a lack of gap minimization occurs with a greater frequency than has generally been reported in the Conversation Analysis literature. These interactions differ from those that have provided the data for many Conversation Analysis studies, in which the speakers are at dinner parties, are on the phone, are in animated groups engaged in lively discussion.

Discussion

In the Stivers et al. (2009) study of turn-taking practices across 10 languages, key findings were that there is a general avoidance of simultaneous talk and a minimization of gaps between turns. However, they did find a variation in the length of average gaps of 0.25 s across these languages, leading them to conclude that the fundamental turn taking mechanisms are universal, with differences between languages being only quantitative. While the methods used for our study and its primary focus are different from theirs, and thus preclude any direct comparison, our findings do broadly support those of Stivers et al. Using the same corpus as for the current study, Mushin and Gardner (2009) noted that approximately 50% of silences in the Garrwa conversations were over 0.9 s, and Gardner (2010), also working with the same corpus, but focusing only on question-answer turns ($N = 62$), found the average gap between question and answer to be 0.75 s, about half as long again as the longest average silences found in a similar environment for any language in the Stivers et al. study[15]. These findings provide further support for the claim that there is some cultural variation in the timing of responses and of next turns generally, but some caution needs to be expressed in making this claim.

In his responses to Stivers and Rossano (2010), on why there may be delays longer than predicted by Sacks et al. (1974), or even no response at all, Schegloff (2010) makes the point that participants in these conversations may be in "continuing states of incipient talk" (Schegloff and Sacks, 1973). Couper-Kuhlen (2010) makes a similar point, reminding us of Goffman's distinction between "focused" and "non-focused" gatherings, with the latter displaying a lack of "tightly organized exchange of doings," and thus perhaps less urgency to produce second pairs parts or, one might add, less urgency to avoid extended silences between turns. Such situations might include "members of a household in their living rooms, employees who share an office, passengers together in an automobile" (Schegloff and Sacks, 1973, pp. 324–325), or, one might further add, old ladies sitting for hours on the porch of a house in the heat of the day. Indeed, the examples from white Australian couples' conversation, where at times similarly long gaps between turns to those in the Garrwa conversations were found, provide some support for Schegloff's and Couper-Kuhlen's observations. In reference to some of the long silences in her data, Jefferson (1989) muses that there may be a "'relaxation" of certain "rules" among intimates' (p. 192). She is referring here specifically to a speaker completing another speaker's turn, where that other speaker is a spouse or sibling. Something similar may be going on in these Garrwa conversations, namely a relaxation of rules amongst intimates where gaps are not minimized. The women in our data grew up together, so are more like family than close friends (in fact, Hilda and Daphne are sisters-in-law); they have lived in small communities in close proximity to each other for many years. Long silences between turns—gaps that do not transform into lapses—are not only tolerated, but are common. In addition, these women have little to do. They sit around for hours at a time, passing the time of day in conversation. Conversations involving intimates in familiar surroundings, with a lack of pressure to talk may in fact be at least as pertinent as cultural difference in accounting for expanded transition spaces.

More specific and local reasons for delays in responding, as some of the examples in the current study suggest, may be "disengagement or lack of attention" (Levinson, 2010), the "activity-occupied withdrawal" that Goodwin (1981) notes, such as brushing one's hair, or seeking something in a handbag. Our analyses also point to the possibility that there is less urgency to respond if a next speaker has not been selected than if they have. Further, but generally less amenable to verification, if a next speaker lacks the knowledge to respond, it is likely that a response may be delayed, or a non-complying response may transpire, or there may be no response at all. Such local factors may then be in play with the wider overall structure of the encounters noted above, and when these local conditions occur during a "non-focused gathering," the frequency of longer gaps between turns may increase.

We do not feel confident to claim that these longer gaps can be explained by culture or ethnicity, even if there is growing evidence that there may be a greater tolerance (or at least occurrence) of longer gaps in some cultures or language groups than others. If gaps are indeed more frequent in some cultures than others, such as Garrwa (and other Indigenous Australian languages) or ≠Ākhoe Hai||om (Hoymann, 2010), spoken in Namibia, it may be that the more traditional life style in very remote areas with relatively little contact to the modern, industrial world does not fully explain the slower pace of conversation, but rather that in such traditional or semi-traditional societies or communities, the people are more likely to live their lives at a slower tempo, in more "non-focused gatherings," and less "tightly organized exchanges of doings." In the conversations we have examined, the old women sit around for hours. They have few appointments to meet, nowhere much to go, little pressure to do anything. They are in familiar surroundings, where they live in close proximity to and know everyone else, as in an extended family. But also they are sitting outside, where people are passing, there are things to watch and notice. Much of the time they are loosely engaged with each other. Life's pace is slow. Conversation is slow. Nevertheless, when the occasion demands, they are perfectly capable of fast-paced conversation, and can provide responses with no gap and no overlap.

As Schegloff (2000) has noted,

> nothing special rests on the "one-at-a-time" proposal. Should a compelling demonstration of a different way of organizing participation in conversation be provided, it would allow us to seek a more general account that could subsume both one-at-a-time and its alternative(s) as special cases
> (p. 47).

The materials presented here in many ways suggest no more than was already accounted for in Sacks et al. (1974) under rule 1c, and some subsequent notes on this rule on how discontinuous talk emerges. If no speaker selects next speaker, no potential next

[15] Gardner (2010) also found that 45% of questions were never answered, which is higher than any of the ten languages investigated by Stivers et al. (2009), which was Korean at 36%.

speaker self-selects, and current speaker chooses not to continue, then a gap develops. Such gaps may occur more or less regularly, and develop into shorter or longer gaps, and extend further or less far toward lapses. In some conversations, with some participants, in some low-pressure situations, and perhaps even in some cultures, it may be the case that the option not to continue under 1c is exercised more regularly, and that once the transition space has passed, the floor opens up to anyone to self-select as next speaker. This could even be posited as a "lowest order rule" of turn-taking: if current speaker chooses not to continue speaking under rule 1c, then after closure of the regular transition space, any speaker may self-select at any time, first speaker to self-select gaining rights to speak. Where the option not to continue speaking is regularly exercised, then the gross observation for "one-at-a-time" will not hold as a recurrent feature of such talk in such circumstances.

There is one final point to be made. We have found that these Garrwa speakers are quite capable of distributing their conversational turns in just the way that Sacks et al. (1974) described, with next speaker selection, self-selection if a next speaker hasn't been selected, and continuation by the current speaker if no other speaker has self selected. The turns in these Garrwa conversations are constructed in units that are identifiable as TCUs, and there is orientation to TRPs, albeit often extended ones. Apart from Sidnell's (2001) study of Caribbean Creole English and Tanaka's (2000) study of Japanese, there have been few studies about the rules of turn-taking of languages other than English, particularly of languages (and cultures) very different from English such as Garrwa, that have shown that the fundamental rules of turn-taking are followed as Sacks, Schegloff, and Jefferson describe. Yet it is through studies such as these that we are able to enrich our understanding of what is fundamental about human social interaction.

Acknowledgments

We would like to thank the Garrwa people who contributed to this project, both in allowing us to record them, and in working with us on the transcriptions. Funding support for this project came from Australian Institute of Aboriginal and Torres Strait Islander Studies grant GO5/7053 Garrwa language project and the University of Queensland. Ethical clearance for the collection of audio and video recordings was given by the University of Queensland Behavioral and Social Sciences Ethical Review Committee (2006000745 Garrwa Language Project).

References

Audacity Team. (2013). *Audacity (R): Free Audio Editor and Recorder [Computer program]. Version 2.0.5.0.* Available online at: http://audacity.sourceforge.net/ (Accessed October 19, 2013).

Blythe, J. (2009). "Prosodic person reference in Murriny Patha reported interaction," in *Where Prosody Meets Interaction,* eds D. Barth-Weingarten, N. Dehe, and A. Wichmann (Bingley: Emerald), 23–52.

Blythe, J. (2013). Preference organization driving structuration: evidence from Australian Aboriginal interaction for pragmatically motivated grammaticalization. *Language* 89, 883–919. doi: 10.1353/lan.2013.0057

Couper-Kuhlen, E. (2010). Commentary on Stivers and Rossano:"mobilizing response." *Res. Lang. Soc. Interact.* 43, 32–37. doi: 10.1080/08351810903471282

Davidson, J. (1984). "Subsequent versions of invitations, offers, requests, and proposals dealing with potential or actual rejection," in *Structures of Social Action,* eds J. Maxwell Atkinson and J. Heritage (Cambridge: CUP), 102–128.

Demosthenous, H. (2008). *Turn-Taking in Hypnosis Session*. Ph.D. thesis, Griffith University, Brisbane.

Eades, D. (2000). I don't think it's an answer to the question: silencing Aboriginal witnesses in court. *Lang. Soc.* 29, 161–195. doi: 10.1017/S0047404500002013

Eades, D. (2007). "Understanding Aboriginal silence in legal contexts," in *Handbook of Intercultural Communication,* eds H. Kotthoff and H. Spencer-Oatey (Berlin: Mouton de Gruyter), 285–301.

Ford, C., and Thompson, S. (1996). "Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns," in *Interaction and Grammar,* eds E. Ochs, E. Schegloff, and S. Thompson (Cambridge: CUP), 134–184.

Fox, B. (2007). Principles in shaping grammatical practices: an exploration. *Discourse Stud.* 9, 299–318. doi: 10.1177/1461445607076201

Fox, B. A., and Thompson, S. A. (2010). Responses to wh-questions in English conversation. *Res. Lang. Soc. Interact.* 43, 133–156. doi: 10.1080/08351811003751680

Garde, M. (2003). *Social Deixis in Bininj Gun-wok*. Unpublished Ph.D. thesis, University of Queensland.

Gardner, R. (2001). *When Listeners Talk: Response Tokens and Recipient Stance. Pragmatics and Beyond Series*. Amsterdam: Benjamins.

Gardner, R. (2007). "Broken starts: bricolage in turn starts in second language talk," in *Language Learning Teaching as Social Interaction*, eds Z. Hua, P. Seedhouse, L. Wei, and V. Cook (London: Palgrave), 58–71.

Gardner, R. (2010). Question and answer sequences in Garrwa talk. *Aust. J. Linguist.* 30, 423–445. doi: 10.1080/07268602.2010.518554

Gardner, R., and Mushin, I. (2007). Post-start-up overlap and disattentiveness in talk in a Garrwa community. *Aust. Rev. Appl. Linguist.* 30, 35.1–35.14. doi: 10.2104/aral0735

Goodwin, C. (1981). *Conversational Organization*. New York, NY: Academic Press.

Goodwin, C. (1994). Professional vision. *Am. Anthropol.* 96, 606–633.

Goodwin, C. (1995a). Seeing in depth. *Soc. Stud. Sci.* 25, 237–274.

Goodwin, C. (1995b). Co-constructing meaning in conversations with an aphasic man. *Res. Lang. Soc. Interact.* 28, 233–260.

Goodwin, C., and Goodwin, M. (1987). Concurrent operations on talk: notes on the interactive organization of assesments. *Pap. Pragmatics* 1, 1–54.

Hayashi, M. (2003). Language and the body as resources for collaborative action: a study of word searches in Japanese conversation. *Res. Lang. Soc. Interact.* 36, 109–141. doi: 10.1207/S15327973RLSI3602_2

Hoymann, G. (2010). Questions and responses in ≠Ākhoe Hai||om. *J. Pragmatics* 42, 2726–2740. doi: 10.1016/j.pragma.2010.04.008

Jefferson, G. (1984). "Notes on some orderliness of overlap onset," in *Discourse Analysis and Natural Rhetoric*, eds V. D'Urso and P. Leonardi (Padua: Cleup Editore), 11–38.

Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Hum. Stud.* 9, 153–183.

Jefferson, G. (1989). "Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation," in *Conversation: An Interdisciplinary Perspective,* eds D. Roger and P. Bull (Clevedon: Multilingual Matters), 166–196.

Kendrick, K. (2015). The intersection of turn-taking and repair: the timing of other-initiation of repair in conversation. *Front. Psychol.* 6:250. doi: 10.3389/fpsyg.2015.00250

Kendrick, K., and Torreira, F. (2014). The timing and construction of preference: a quantitative study. *Discourse Process*. doi: 10.1080/0163853X.2014.955997. [Epub ahead of print].

Kita, S., and Ide, S. (2007). Nodding, aizuchi, and final particles in Japanese conversation: how conversation reflects the ideology of communication and social relationships. *J. Pragmatics* 39, 1242–1254. doi: 10.1016/j.pragma.2007.02.009

Levinson, S. (1983). *Pragmatics*. Cambridge: CUP.
Levinson, S. (2010). Questions and responses in Yélî Dnye, the Papuan language of Rossel Island. *J. Pragmatics* 42, 2741–2755. doi: 10.1016/j.pragma.2010.04.009
Moerman, M. (1988). *Talking Culture: Ethnography and Conversation Analysis*. Pennsylvania, PA: University of Pennsylvania Press.
Mushin, I., and Gardner, R. (2009). Silence is talk: conversational silence in Australian Aboriginal talk-in-interaction. *J. Pragmatics* 40, 2033–2052. doi: 10.1016/j.pragma.2008.11.004
Mushin, I., and Gardner, R. (2011). "Turn management in Garrwa mixed-language conversations," in *Language and Social Identity in Indigenous Commmunities: Papers Presented to Michael Walsh on the Occasion of his Retirement,* eds B. Baker, M. Harvey, I. Mushin, and R. Gardner (Canberra, ACT: Pacific Linguistics).
Philips, S. (1976). Some sources of cultural variability in the regulation of talk. *Lang. Soc.* 5, 81–95.
Philips, S. (1983). *The Invisible Culture: Communication in Classroom and Community on the Warm Springs Indian Reservations*. Prospect Heights, NY: Waveland.
Pomerantz, A. (1984a). "Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes," in *Structures of Social Action,* eds J. Maxwell Atkinson and J. Heritage (Cambridge: CUP), 57–101.
Pomerantz, A. (1984b). "Pursuing a response," in *Structures of Social Action,* eds J. Maxwell Atkinson and J. Heritage (Cambridge: CUP), 152–163.
Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735.
Schegloff, E. (1996). "Turn organization: one intersection of grammar and interaction," in *Interaction and Grammar,* eds E. Ochs, E. Schegloff, and S. Thompson (Cambridge: CUP), 52–133.
Schegloff, E. (1998). Body torque. *Soc. Res.* 65, 535–596.
Schegloff, E. (2000). Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 1, 1–63.
Schegloff, E. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
Schegloff, E. (2010). Commentary on Stivers and Rossano:"mobilizing response." *Res. Lang. Soc. Interact.* 43, 38–48. doi: 10.1080/08351810903471282
Schegloff, E. A., and Sacks, H. (1973). Opening up closings. *Semiotica* 8, 289–327.
Scollon, R., and Scollon, S. (1981). *Narrative, Literacy and Faith in Interethnic Communication*. Norwood, NJ: Ablex.
Sidnell, J., and Stivers, T. (eds.). (2012). *The Handbook of Conversation Analysis,* Vol. 121. Chichester: John Wiley & Sons.
Sidnell, J. (2001). Conversational turn-taking in a Caribbean English Creole. *J. Pragmatics* 33, 1263–1290. doi: 10.1016/S0378-2166(00)00062-X
Sidnell, J. (2011). *Conversation Analysis: An Introduction,* Vol. 45. Chichester: John Wiley & Sons.
Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106
Stivers, T., and Robinson, J. (2006). A preference for progressivity in interaction. *Lang. Soc.* 35, 367–392. doi: 10.1017/S0047404506060179
Stivers, T., and Rossano, F. (2010). Mobilizing response. *Res. Lang. Soc. Interact.* 43, 3–31. doi: 10.1080/08351810903471258
Streek, J. (1993). Gesture as communication 1: its coordination with gaze and speech. *Commun. Monogr.* 60, 275–299.
Tanaka, H. (2000). *Turn-Taking in Japanese Conversation: A Study in Grammar and Interaction,* Vol. 56. Amsterdam: John Benjamins.
Walsh, M. (1991). Conversational styles and intercultural communication: an example from northern Australia. *Aust. J. Commun.* 18, 1–12.
Wells, B., and Macfarlane, S. (1998). Prosody as an interactional resource: turn-projection and overlap. *Lang. Speech* 41, 265–294.
Wilson, T., and Zimmerman, D. (1986). The structure of silence between turns in two-party conversation. *Discourse Process.* 9, 375–390.
Wong, J. (2000). Delayed next turn repair initiation in native/non-native speaker English conversation. *Appl. Linguist.* 21, 244–267. doi: 10.1093/applin/21.2.244

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

PERSPECTIVE
published: 14 April 2015
doi: 10.3389/fpsyg.2015.00450

Experience sharing, emotional reciprocity, and turn-taking

Melisa Stevanovic and Anssi Peräkylä*

Finnish Centre of Excellence in Intersubjectivity in Interaction, Department of Social Research, University of Helsinki, Helsinki, Finland

In this perspective article, we consider the relationship between experience sharing and turn-taking. There is much evidence suggesting that human social interaction is permeated by two temporal organizations: (1) the sequential framework of turn-taking and (2) the concurrent framework of emotional reciprocity. From this perspective, we introduce two alternative hypotheses about how the relationship between experience sharing and turn-taking could be viewed. According to the first hypothesis, the home environment of experience sharing is in the concurrent framework of emotional reciprocity, while the motivation to share experiences is in tension with the sequential framework of turn-taking. According to the second hypothesis, people's inclination to coordinate their actions in terms of turn-taking is motivated precisely by their propensity to share experiences. We consider theoretical and empirical ideas in favor of both of these hypotheses and discuss their implications for future research.

Keywords: turn-taking, emotion, experience sharing, emotional contagion, conversation analysis

OPEN ACCESS

Edited by:
Kobin H. Kendrick, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Beatrice S. Reed, University of York, UK
Alexa Hepburn, Loughborough University, UK

****Correspondence:***
Melisa Stevanovic, Finnish Centre of Excellence in Intersubjectivity in Interaction, Department of Social Research, University of Helsinki, P.O.Box 4, 00014 University of Helsinki, Finland
melisa.stevanovic@helsinki.fi

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *19 November 2014*
Accepted: *30 March 2015*
Published: *14 April 2015*

Citation:
Stevanovic M and Peräkylä A (2015) Experience sharing, emotional reciprocity, and turn-taking. Front. Psychol. 6:450. doi: 10.3389/fpsyg.2015.00450

Sequentiality and Concurrency in Human Social Interaction

In recent years, there has been an increasing number of conversation analytic studies explicating the social organization of, and the highly ordered interactional tasks performed by, emotional expressions in social interaction (e.g., Peräkylä and Sorjonen, 2012). This perspective paper seeks to further this line of research by considering how expressions in the service of what we call experience sharing are embedded in the most primordial temporal organizations of interaction.

Many naturally occurring interactions call for individuals to coordinate their actions in terms of turn-taking. This happens especially in the context of language use: the principle of one participant talking at a time (Sacks et al., 1974) allows humans to communicate complex thoughts and intentions. In conversation, social actions (e.g., proposals, offers, and invitations) and their responses (e.g., acceptances and rejections) are organized in terms of successive turns at talk. As pointed out by Schegloff (1988, pp. 98–99), turn-taking enables humans to pursue stable trajectories of action and responsive action. This arrangement will be referred to as the *sequential framework of turn-taking*. It denotes not only the temporal but also the conditional relationship between participants' interactional moves.

In addition to the sequential framework of turn-taking, human interactive conduct is permeated by another temporo-conditional arrangement—something that we call the *concurrent framework of emotional reciprocity*. The literature on "emotional contagion" suggests that humans have an automatic tendency to mimick other people's non-verbal emotional expressions, which affects the emotional experience of the mimicking person (Hatfield et al., 1993, 1994; Dimberg, 2007). This happens as a result of afferent feedback generated by elementary motor mimicry, which produces a simultaneous emotional match independently of people's cognitive abilities to understand what is

going on in the mind of the other (Carr et al., 2003; Leslie et al., 2004; Barresi and Moore, 2008; Decety and Meyer, 2008). From this perspective, two participants' interactional moves are, by definition, connected by a causal relationship (one participant produces an expression first and the other acts in response to him/her), but the actions/expressions take place in a shared time. Overlap of expressions is regular. Regarding the opportunity for expression, the participants are positioned symmetrically, which is reflected in the frequent concurrency of the participants' interactional moves.

There are different motives that lead humans to interact with each other. In addition to their instrumental communicative goals, humans are also motivated to share experiences about events and things in the environment with their significant others and to become "swept along" by them (Feinman, 1982; Striano and Rochat, 1999; Tomasello, 1999; Hobson and Hobson, 2008; Rochat et al., 2009). This motivation has even been regarded as the force that has driven the evolution of language (Locke, 1996, 2002; Dunbar, 1997). But how does experience sharing relate to the sequential framework of turn-taking, where language use regularly takes place? Could it be that experience sharing is "at home" in the concurrent framework of emotional reciprocity, while being in tension with turn-taking? Or, do humans cast their experience sharing into the system of turn-taking precisely because it supports experience sharing? In what is to come, we will discuss these two alternative hypotheses one after another, with the aim of paving the way for future empirical research on the topic.

Hypothesis I: Experience Sharing is in Tension with the Sequential Framework of Turn-Taking

Our first hypothesis suggests that the home environment of experience sharing is in the concurrent framework of emotional reciprocity, while there is a tension between experience sharing and the sequential framework of turn-taking.

The first pieces of support for this hypothesis come from the developmental psychological research literature, which suggests there to be automatic resonance processes that allow humans, right from the outset, to bridge their own and others' experiences. With reference to neonatal imitation, Meltzoff and Brooks (2001) have argued that, in reproducing the behavior of others, infants automatically perceive others as "like me" and thus begin to develop a sense of social connectedness, mutual recognition, and shared experience. In other words, experience sharing has been suggested to emerge "in the guise of emotional contagion" (Brinck, 2008).

Another type of support for our first hypothesis comes from adult interaction, and is provided by the temporal organization of the instances of language use associated with experience sharing. While language use may be anchored in the organization of turn-taking (see e.g., Schegloff, 1996, 2006), still, there is much work suggesting that the moments of experience sharing may be exceptional in this respect (Coates, 1994; Lerner, 2002; Pillet-Shore, 2012; Vatanen, 2014). In conversation analysis, a classic example is provided by Goodwin and Goodwin (1987), who described the sharing of affective stances in the form of concurrent agreeing assessments. Their example involves two conversationalists praising something a mutual friend has baked. One of them says: "Jeff made an asparagus pie, it was so good." In overlap with the first speaker's "so," the co-interactant launches an assessment: "I love it." Hence, it appears that, in the moments of experience sharing, the concurrent framework of emotional reciprocity colonizes the organization of spoken interaction, leading to the momentary relaxation of turn-taking rules. Also studies within the domain of mother-infant interaction have shown that positive affective expressions tend to coincide with simultaneous vocalizations (Stern et al., 1975; Beebe et al., 1979).

Besides overlapping talk, there are also other ways in which experience sharing, as it were, "surpasses" turn-taking. Face is central here. Speakers may use their facial displays to mark a transition from affectively neutral talk to emotional experience sharing in the middle of their ongoing turns at talk (cf. Iwasaki, 2011). Likewise, a recipient may display an emotional stance toward an actional or a topical element in a speaker's ongoing turn at talk, thereby inviting the speaker to redirect her utterance production (Kaukomaa et al., in press). Detailed considerations of parallel uses of words and facial displays in moments of experience sharing are thus particularly intriguing.

From this perspective, let us consider Example 1 (taken from Peräkylä and Ruusuvuori, 2006), where one participant's telling (about a dress code that a mutual friend working in a newspaper needed to conform to) is followed by a shared amusement of both of the participants. Here, the onset and the completion of the smiles take place with one party doing the first move and the other party following. However, between these sequentially organized boundary moves, the participants maintain simultaneous smiles, embodying the sharing of experiences over a lengthy period of time.

During the most part of the first speaker's (B's) telling (lines 1–6) both participants are looking down with straight faces (Frame 1). At the end of the telling, however, they establish mutual gaze and the teller (B) adopts a smiling face (Frame 2). After a gap (line 7), the recipient (A) reciprocates the smile and raises her brows (Frame 3), simultaneously producing an animated verbal response of "ritualized disbelief" (Heritage, 1984, p. 339; line 8). While the interaction has thus far abided to the sequential framework of turn-taking, now, as soon as the mutual smile has been established, the participants' facial conduct gets detached from the sequential organization of turn-taking as the concurrent framework of emotional reciprocity breaks through. During a moment of "heightened emotive involvement" (Selting, 1994), they maintain their smiles and mutual gaze over a number of turns conveying ostensibly different actions (assessment, line 10; joking advice, line 11; Frame 4). The temporality of the mutual smile goes beyond the turn-taking organization. After this, the participants, however, break their mutual gaze (Frame 5), which is followed by their smiles becoming less intensive. One participant (A) adopts a straight face at the onset of her question (lines 13–14), while the other (B) does the same during her answer to the question (line 15; Frame 6). Thereby, they indicate that the heightened moment of experience sharing is over. While the primary modality of expression surpassing the

Example 1

01 B: t Savon Sanomissa oli (0.7) ei Savon
tch in Savon Sanomat they had (0.7) not in Savon

02 Sanomissa ku (1.6) Sata#kunnan Kansassa#.
Sanomat but in (1.1)Sata#kunnan Kansa#.

03 (0.4)

04 A: Mm//m,
Fr1

Fr. 1:
Straight faces

Fr. 2:
Speaker smile at utterance completion

05 (0.4)

06 B: Oli kehotettu Samia laittaa #suorat hou°sut°<#. //
They had asked Sami to wear #proper trou°sers<°.#//
Fr2

07 (0.4)

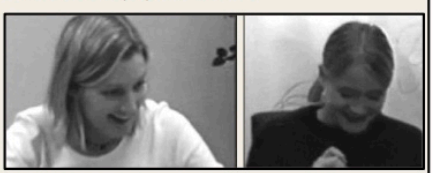
08 A: M:i(h)tä//:.
W:(h)ha:t//:.
Fr3

09 (0.7)

Frs. 3-5:
Smile reciprocation and mutual smile maintenance across turns at talk

10 A: Ei kauh//[eet(h)a(h).]
No that's horr//[ibl(h)e(h).]
Fr4

11 B: [Et älä si]nne ainakaan hae //t(h)olla.
[So don't app]ly there in any case // with th(h)at one.
Fr 5



12 A: .hhh Ohoh.
.hhh oh no.

13 A: .hhh Minkä paikan lehti siis Satakunnan
.hhh at which town is that paper so Satakunnan

14 Sa[nomat.
Sa[nomat

15 B: [Pori.//
[At Pori.//
Fr6

Fr 6:
Return to straight faces / affective neutrality

turn-taking rules in Example 1 was facial expression, and the organization of talk followed turn-taking rules, the overlapping talk (lines 10–11) during the experience sharing should also be acknowledged.

The idea of experience sharing being in tension with the sequential temporality of turn-taking reverberates with certain recent suggestions presented in the conversation analytic literature: Heritage (2011) has described the epistemic dilemmas associated with the "empathic moments in interaction" and Enfield (2011) has suggested that it is turn-taking with its inherent asymmetries that helps to account for the existence of such dilemmas. But even if experience sharing would be in tension with turn-taking, this seems not to be the case for instrumental communication. In her study on university subcommittee meetings, Edelsky (1981) observed that those participants who otherwise made frequent use of overlapping talk still abided to the canonical turn-taking system when their talk was oriented toward the official business of the meeting. The same phenomenon seems to apply also for those emotional expressions that are used to carry out different kinds of instrumental communicative tasks. For example, Heath (1989) showed that, when patients in medical consultations tried to legitimize medical attention to their ailment through cries of pain, these expressions of suffering abided to the organization of turn-taking.

The points detailed above suggest that the home environment for experience sharing might be in the concurrent framework of emotional reciprocity, while experience sharing (unlike instrumental communicative goals) may call for the participants to depart from the sequential framework of turn-taking.

Hypothesis II: The Sequential Framework of Turn-Taking Serves Experience Sharing

While our first hypothesis questions the relevance of turn-taking for experience sharing, our second hypothesis represents just the opposite view: it suggests that the sequential framework of turn-taking is *not* in tension with experience sharing but, instead, serves it.

Previously, we highlighted the significance of the automatic resonance processes for experience sharing. However, it has been pointed out that the mere reproduction of other people's behaviors represents a closed loop system: it reflects what is already out there (Rochat and Passos-Ferreira, 2008). So, for people really to relate and share each other's experiences more fully, these automatic resonance processes need to be embedded in an open system of *contingent* emotional reciprocity. It can thus be argued that contingent emotional reciprocity, and thus also experience sharing, can be facilitated by the sequential framework of turn-taking.

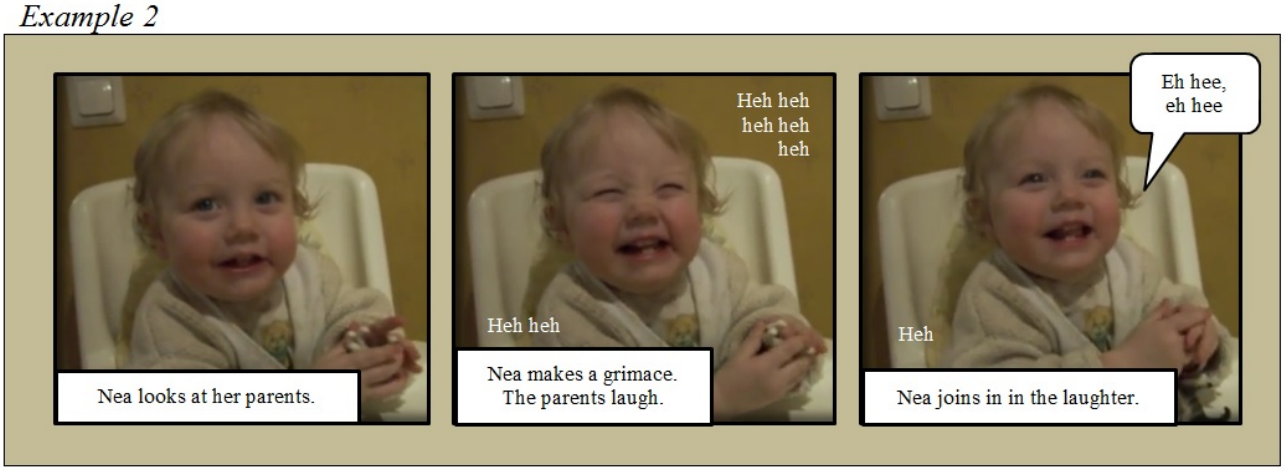
In human ontogeny, the first instantiations of contingent emotional reciprocity appear in the context of alternation between approach and withdrawal tendencies (Beebe and Stern, 1977; Hietanen et al., 2008). Infants have been shown occasionally visually to disengage from their interactive partners and then return to the engagement. Often, this happens in a highly coordinated fashion—that is, when one partner moves from a less engaged phase to a more engaged phase, or vice versa, the other partner responds with a corresponding change in the same direction. Still, these changes are likely to take place within certain time-lags (Cohn and Tronick, 1987). This gives the interaction a sense of one person making a bid of engagement and another person responding to that bid in a positive way.

In the subsequent development of the human infant, the experiences of contingent emotional reciprocity get more nuanced. Already from 2 months on, infants and their caretakers start to look and listen to each other more carefully; producing vocal, facial and gestural responses elicited by the expressed feelings and interests of their interaction partners (Spitz and Wolf, 1946; Trevarthen, 1979; Stern, 1985; Cohn and Elmore, 1988; Rochat and Passos-Ferreira, 2008). During these monitoring processes, the infants gradually develop expectations for how the interaction is likely to proceed; for example, they learn to expect that, following an emotional bid on their part, be it via a smile, gaze, or frown, the other will respond in return (Sagi and Hoffman, 1976; Meltzoff and Moore, 1977; Wolff, 1987; Sroufe, 1996; Rochat, 2001). Compared to the automatic resonance processes, the reliance on social expectations is risky but, when successful, likely to result in a powerful experience of shared emotion (Rochat et al., 2009).

From this perspective, let us consider Example 2, where an 11-month-old girl, Nea, prompts a prominent instance of experience sharing. First, she looks at her parents, assuring their attention. Thereafter, she puts a funny grimace on her face, thus prompting her parents' to laugh heartfeltly. Finally, she joins in the laughter, expressing a high level of positive arousal. The sharing of experience is organized in successive turns. This turn-taking organization of sharing builds on the child's capacity to anticipate her significant others' reactions to her behavior, while there is a genuine possibility that the parents will not behave as expected.

Example 2 demonstrates how a turn-taking structure, where expressions are organized as distinct temporal units that follow each other, provides for experiences of contingent emotional reciprocity. By making a grimace, Nea, took on trust that her parents would recognize what she was up to; in their "next turn" (laughter), the parents displayed that they indeed did; and, the correctness of the parents' interpretation was confimed by Nea, through her final laughter. Here, turn-taking enables the participants to establish meaningful linkages between their behaviors. While each behavior occupies a forward-looking status, raising expectations for what is to happen next (on conditional relevance,

Example 2



see Schegloff, 2007, p. 20), each response to such behaviors occupies a backward-looking status as something that was invoked by what just occurred (on the next-turn proof procedure, see Sacks et al., 1974, pp. 728–729). It is thus the dynamic interplay between expectations and their overflowing satisfaction that provides a scaffolding for a possible escalation of emotion displays (each participant endorsing the affective aspect of their co-participants' turns), which may generate particularly intensive instances of experience sharing. Also the conversation analytic research on emotional expression provides support for our second hypothesis. This line of research has shown that the participants may refrain from the immediate reciprocation of contagious emotional expressions, such as laughter and crying (Hepburn and Potter, 2012; Shaw et al., 2013). Instead, these expressions seem to be regulated by the turn-by-turn sequential organization of interaction. For example, the recipients of complaint stories (Couper-Kuhlen, 2012) and news deliveries (Maynard and Freese, 2012) have been shown to produce their emotional responses at the completion of the news delivery or narrative, rather than immediately after the tellers' emotional displays. Likewise, surprise tokens such as *wow*, *gosh*, *oh my good*, *ooh*, *phew* have been described as interactionally organized performances, interactional achievements, instead of automatic emotional eruptions (Wilkinson and Kitzinger, 2006; see also Heath et al., 2012). It could be precisely the sensitivity of emotional expression to the sequential framework of turn-taking that enables the participants to display their willingness to grasp—as fully as possible—the particular experience that their co-participants are about to share.

Finally, we may turn the table around and ask whether experience sharing can also serve turn-taking. This possibility appears relevant with reference to the Goffmanian idea of the ubiquitous insecurity of people in social interaction—a theme that runs all through his work (see e.g., Goffman, 1955; Rawls, 1987). In social interaction, by every turn that a person takes, s/he claims being worthy of other's attention and calls for others to recognize this claim (Goffman, 1955, pp. 9–10), while there is always a possibility that this claim will *not* be recognized by others (Stevanovic, 2015; Peräkylä, in press). The emotionally secure framework of experience sharing with significant others, however, provides an embarrassment-free site for the practicing of making such claims and for the acquisition of the more asymmetric responsibilities that the sequential framework of turn-taking enforces the participants to assume.

In sum, the coordination of actions in terms of turn-taking enables forms of experience sharing that the concurrent organization could not afford. This suggests that one motivation underlying the human propensity to coordinate their actions in terms of turn-taking could be the human propensity to share experiences. Even more, there might be a bidirectional linkage of enhancement between these two propensities.

Consequences for the Study of Turn-Taking

We have now considered two alternative hypotheses about how the relationship between experience sharing and turn-taking could be viewed. What consequences would these imply for the study of turn-taking?

According to our first hypothesis, the home environment of experience sharing is in the concurrent framework of emotional reciprocity, while there is a tension between experience sharing and the sequential framework of turn-taking. If this hypothesis is valid, then there must be evidence that, not only does a lesser amount of experience sharing lead to a greater amount of turn-taking, and vice versa, but also, that turn-taking obstructs experience sharing. Furthermore, if turn-taking is there to facilitate instrumental communication (instead of experience sharing), then we would expect that there would be a tension between overlapping vocalizations and effective instrumental communication. While there is some evidence for the tension between turn-taking and experience sharing (Enfield, 2011; Heritage, 2011; Vatanen, 2014), the potentially problematic combination of overlapping vocalizations and effective instrumental communication has rather been taken for granted than really unpacked through empirical investigation (cf. Stevanovic and Frick, 2014). One further challenge from the perspective of this hypothesis would be to account for the occurrences of overlapping talk serving instrumental purposes.

According to the second hypothesis, turn-taking is in the service of experience sharing. In allowing the participants to engage in increasingly complex forms of joint action, and, hence, in ever more exciting shared experiences, there seems to be no tension between turn-taking and experience sharing. From this perspective, there seems to be a developmental continuity between the early infant-caretaker interactions governed by emotional reciprocity and the later, more complex forms of social interaction. If this hypothesis is valid, then there must be evidence that the instances of experience sharing cast in the sequential framework of turn-taking can, in principle, reach at least the same level of intensity as those occurring in the context of overlapping talk. Moreover, one would need to show that occurrences of experience sharing and instrumental goal-pursuit would be relatively evenly distributed between the instances of overlapping talk and talk abiding to the norms of turn-taking.

No matter which of the above hypothese is more valid than the other, the relationship between experience sharing and turn-taking is something worth further investigation—something that this paper has sought to highlight.

Acknowledgments

This work was supported by the Academy of Finland (Decision No. 274735).

References

Barresi, J., and Moore, C. (2008). "The neuroscience of social understanding," in *The Shared Mind: Perspectives on Intersubjectivity*, eds J. Zlatev, T. P. Racine, C. Sinha, and E. Itkonen (Amsterdam: Benjamins), 39–66. doi: 10.1075/celcr.12.05bar

Beebe, B., and Stern, D. (1977). "Engagement-disengagement and early object experiences," in *Communicative Structures and Psychic Structures*, eds N. Freedman and S. Grand (New York: Plenum), 35–55. doi: 10.1007/978-1-4757-0492-1_3

Beebe, B., Stern, D., and Jaffe, J. (1979). "The kinesic rhythm of mother-infant interactions," in *Of Speech and Time: Temporal Patterns in Interpersonal Contexts*,

eds A. Siegman and S. Feldstein (Hillsdale, NJ: Lawrence Erlbaum Associates), 23–34.

Brinck, I. (2008). "The role of intersubjectivity in the development of intentional communication," in *The Shared Mind: Perspectives on Intersubjectivity*, eds J. Zlatev, T. P. Racine, C. Sinha, and E. Itkonen (Amsterdam: Benjamins), 115–140. doi: 10.1075/celcr.12.08bri

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., and Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5497–5502. doi: 10.1073/pnas.0935845100

Coates, J. (1994). "No gap, lots of overlap: turn-taking patterns in the talk of women friends," in *Researching Language and Literacy in Social Context*, eds D. Graddol, J. Maybin, and B. Stierer (Clevedon: Multilingual Matters), 177–192.

Cohn, J. E., and Elmore, M. (1988). Effect of contingent changes in mothers' affective expression on the organization of behavior in 3-month-old infants. *Infant Behav. Dev.* 11, 493–505. doi: 10.1016/0163-6383(88)90008-2

Cohn, J. E., and Tronick, E. Z. (1987). Mother-infant face-to-face interaction: the sequence of dyadic states at 3, 6, and 9 months. *Dev. Psychol.* 23, 68–77. doi: 10.1037/0012-1649.23.1.68

Couper-Kuhlen, E. (2012). "Exploring affiliation in the reception of conversational complaint stories," in *Emotion in Interaction*, eds A. Peräkylä and M.-L. Sorjonen (Oxford: Oxford University Press), 113–146.

Decety, J., and Meyer, M. (2008). From emotion resonance to empathic understanding: a social developmental neuroscience account. *Dev. Psychopathol.* 20, 1053–1080. doi: 10.1017/S0954579408000503

Dimberg, U. (2007). Facial reactions to facial expressions. *Psychophysiology* 19, 643–647. doi: 10.1111/j.1469-8986.1982.tb02516.x

Dunbar, R. (1997). *Grooming, Gossip, and the Evolution of Language*. Cambridge, MA: Harvard University Press.

Edelsky, C. (1981). Who's got the floor? *Lang. Soc.* 10, 383–421. doi: 10.1017/S004740450000885X

Enfield, N. (2011). "Sources of asymmetry in human interaction: enchrony, status, knowledge and agency," in *The Morality of Knowledge in Conversation*, eds T. Stivers, L. Mondada, and J. Steensig (Cambridge: Cambridge University Press), 285–312.

Feinman, S. (1982). Social referencing in infancy. *Merrill Palmer Q. J. Dev. Psychol.* 28, 445–470.

Goffman, E. (1955). On face-work: an analysis of ritual elements in social interaction. *Psychiatry* 18, 213–231.

Goodwin, C., and Goodwin, M. H. (1987). Concurrent operations on talk: notes on the interactive organisation of assessments. *IPRA Papers Pragmat.* 1, 1–54. doi: 10.1075/iprapip.1.1.01goo

Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1993). Emotional contagion. *Curr. Dir. Psychol. Sci.* 2, 96–99. doi: 10.1111/1467-8721.ep10770953

Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1994). *Emotion Contagion*. New York: Cambride University Press.

Heath, C. (1989). Pain talk: the expression of suffering in the medical consultation. *Soc. Psychol. Q.* 52, 113–125. doi: 10.2307/2786911

Heath, C., vom Lehm, D., Cleverly, J., and Luff, P. (2012). "Revealing surprise: the local ecology and the transposition of action," in *Emotion in Interaction*, eds A. Peräkylä and M.-L. Sorjonen (Oxford: Oxford University Press), 212–234. doi: 10.1093/acprof:oso/9780199730735.003.0010

Hepburn, A., and Potter, J. (2012). "Crying and crying responses," in *Emotion in Interaction*, eds A. Peräkylä and M.-L. Sorjonen (Oxford: Oxford University Press), 194–210.

Heritage, J. (1984). "A change-of-state token and aspects of its sequential placement," in *Structures of Social Action*, eds J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 299–345.

Heritage, J. (2011). "Territories of knowledge, territories of experience: empathic moments in interaction," in *The Morality of Knowledge in Conversation*, eds T. Stivers, L. Mondada, and J. Steensig (Cambridge: Cambridge University Press), 159–183.

Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K., and Ruuhiala, H. J. (2008). Seeing direct and averted gaze activates the approach-avoidance motivational brain systems. *Neuropsychologia* 46, 2423–2430. doi: 10.1016/j.neuropsychologia.2008.02.029

Hobson, P., and Hobson, J. A. (2008). "Engaging, sharing, knowing: some lessons from research in autism," in *The Shared Mind: Perspectives on Intersubjectivity*, eds J. Zlatev, T. P. Racine, C. Sinha, and E. Itkonen (Amsterdam: Benjamins), 67–88. doi: 10.1075/celcr.12.06hob

Iwasaki, S. (2011). "The multimodal mechanics of collaborative unit construction in Japanese conversation," in *Embodied Interaction: Language and the Body in the Material World*, eds J. Streeck, C. Goodwin, and C. LeBaron (Cambridge: Cambridge University Press), 106–120.

Kaukomaa, T., Peräkylä, A., and Ruusuvuori, J. (in press). Facial expressions in collaborative modification of emotional stance. *Res. Lang. Soc. Int.*

Lerner, G. (2002). "Turn-sharing: the choral co-production of talk-in-interaction," in *The Language of Turn and Sequence*, eds C. E. Ford, B. A. Fox, and S. A. Thompson (New York: Oxford University Press), 225–256.

Leslie, K. R., Johnson-Frey, S. H., and Grafton, S. T. (2004). Functional imaging of face and hand imitation: towards a motor theory of empathy. *Neuroimage* 21, 601–607. doi: 10.1016/j.neuroimage.2003.09.038

Locke, J. L. (1996). Why do infants begin to talk? Language as an unintended consequence. *J. Child Lang.* 23, 251–268. doi: 10.1017/S0305000900008783

Locke, J. L. (2002). First communion: the emergence of vocal relationships. *Soc. Dev.* 10, 294–308. doi: 10.1111/1467-9507.00167

Maynard, D. W., and Freese, J. (2012). "Good news, bad news, and affect: practical and temporal 'emotion work' in everyday life," in *Emotion in Interaction*, eds A. Peräkylä and M.-L. Sorjonen (Oxford: Oxford University Press), 92–112. doi: 10.1093/acprof:oso/9780199730735.003.0005

Meltzoff, A. N., and Brooks, R. (2001). "'Like me' as a building block for understanding other minds: bodily acts, attention, and intention," in *Intentions and Intentionality: Foundations of Social Science*, eds B. Malle, L. J. Moses and D. A. Baldwin (Cambridge, MA: MIT Press), 171–191.

Meltzoff, A. N., and Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science* 198, 75–78. doi: 10.1126/science.198.4312.75

Peräkylä, A., and Ruusuvuori, J. (2006). "Facial expression in an assessment," in *Video-analysis: Methodology and Methods*, eds H. Knoblauch, B. Schnettler, J. Raab, and H. G. Soeffner (Frankfurt am Main: Peter Lang), 127–142.

Peräkylä, A., and Sorjonen, M.-L. (eds). (2012). *Emotion in Interaction*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199730735.001.0001

Peräkylä, A. (in press). From narcissism to face work: two views on the self in social interaction. *Amer. J. Soc.*

Pillet-Shore, D. (2012). Greeting: displaying stance through prosodic recipient design. *Res. Lang. Soc. Int.* 45, 375–398. doi: 10.1080/08351813.2012.724994

Rawls, A. W. (1987). The interation order sui generis: goffman's contribution to social theory. *Soc. Theory* 5, 136–149. doi: 10.2307/201935

Rochat, P. (2001). *The Infant's World*. Cambridge, MA: Harvard University Press.

Rochat, P., and Passos-Ferreira, C. (2008). "From imitation to reciprocation and mutual recognition," in *Mirror Neurons System: The Role of Mirroring Processes in Social Cognition*, ed. J. A. Pineda (New York: Humana Press, Springer Science), 191–212.

Rochat, P., Passos-Ferreira, C., and Salem, P. (2009). "Three levels of intersubjectivity in early development," in *Enacting Intersubjectivity: Paving the Way for a Dialogue between Cognitive Science, Social Cognition and Neuroscience*, eds A. Carassa, F. Morganti, and G. Riva (Lugano: Università della Svizzera Italiana), 173–190.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language* 50, 696–735. doi: 10.2307/412243

Sagi, A., and Hoffman, M. L. (1976). Empathic distress in the newborn. *Dev. Psychol.* 12, 175–176. doi: 10.1037/0012-1649.12.2.175

Schegloff, E. A. (1988). "Goffman and the analysis of conversation," in *Erving Goffman: Exploring the Interaction Order*, eds P. Drew and A. Wootton (Cambridge: Polity Press), 89–135.

Schegloff, E. A. (1996). "Turn organization: one direction for inquiry into grammar and interaction," in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 52–133.

Schegloff, E. A. (2006). "Interaction: the infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is situated," in *Roots of Human Sociality: Culture, Cognition and Interaction*, eds N. J. Enfield and S. C. Levinson (Oxford: Berg), 70–96.

Schegloff, E. A. (2007). *Sequence Organization in Interaction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511791208

Selting, M. (1994). Emphatic speech style: with special focus on the prosodic signalling of heightened emotive involvement in conversation. *J. Pragmat.* 22, 375–408. doi: 10.1016/0378-2166(94)90116-3

Shaw, C., Hepburn, A., and Potter, J. (2013). "Having the last laugh: on post-completion laughter particles," in *Studies of Laughter in Interaction*, eds P. Glenn and E. Holt (London: Bloomsbury), 91–106.

Spitz, R. A., and Wolf, K. M. (1946). The smiling response: a contribution to the ontogenesis of social relations. *Genet. Psychol. Monographs* 34, 57–125.

Sroufe, L. A. (1996). *Emotional Development: The Organization of Emotional Life in the Early Years*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511527661

Stern, D. (1985). *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. New York, NY: Basic Books.

Stern, D. N., Jaffe, J., Beebe, B., and Bennett, S. L. (1975). Vocalizing in unison and in alternation: two modes of communication within the mother-infant dyad. *Ann. N. Y. Acad. Sci.* 263, 89–100. doi: 10.1111/j.1749-6632.1975.tb41574.x

Stevanovic, M. (2015). Displays of uncertainty and proximal deontic claims: the case of proposal sequences. *J. Pragmat.* 78, 84–97. doi: 10.1016/j.pragma.2014.12.002

Stevanovic, M., and Frick, M. (2014). Singing in interaction. *Soc. Semiotics* 24, 495–513. doi: 10.1080/10350330.2014.929394

Striano, T., and Rochat, P. (1999). Developmental link between dyadic and triadic social competence in infancy. *Br. J. Dev. Psychol.* 17, 551–562. doi: 10.1348/026151099165474

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Trevarthen, C. (1979). "Communication and cooperation in early infancy: a description of primary intersubjectivity," in *Before Speech: The Beginnings of Human Communication*, ed. M. Bullowa (London: Cambridge University Press), 99–136.

Vatanen, A. (2014). *Responding in Overlap: Agency, Epistemicity and Social Action in Conversation*. Academic dissertation. University of Helsinki, Department of Finnish, Finno-Ugric and Scandinavian Studies.

Wilkinson, S., and Kitzinger C. (2006). Surprise as an interactional achievement: reaction tokens in conversation. *Soc. Psychol. Q.* 69, 150–182. doi: 10.1177/019027250606900203

Wolff, P. (1987). *The Development of Behavioral States and the Expression of Emotions in Early Infancy*. Chicago, IL: The University of Chicago Press.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 March 2015
doi: 10.3389/fpsyg.2015.00268

Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries

Connie de Vos[1]*, *Francisco Torreira*[1] and *Stephen C. Levinson*[1,2]

[1] Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [2] Radboud University, Nijmegen, Netherlands

In spoken interactions, interlocutors carefully plan, and time their utterances, minimizing gaps and overlaps between consecutive turns. Cross-linguistic comparison has indicated that spoken languages vary only minimally in terms of turn-timing, and language acquisition research has shown pre-linguistic vocal turn-taking in the first half year of life. These observations suggest that the turn-taking system may provide a fundamental basis for our linguistic capacities. The question remains, however, to what extent our capacity for rapid turn-taking is determined by modality constraints. The avoidance of overlapping turns could be motivated by the difficulty of hearing and speaking at the same time. If so, turn-taking in sign might show greater toleration for overlap. Alternatively, signed conversations may show a similar distribution of turn-timing as spoken languages, thus avoiding both gaps and overlaps. To address this question we look at turn-timing in question–answer sequences in spontaneous conversations of Sign Language of the Netherlands. The findings indicate that although there is considerable overlap in two or more signers' articulators in conversation, when proper allowance is made for onset preparation, post-utterance retraction and the intentional holding of signs for response, turn-taking latencies in sign look remarkably like those reported for spoken language. This is consistent with the possibility that, at least with regard to responses to questions, speakers and signers follow similar time courses in planning and producing their utterances in on-going conversation. This suggests that turn-taking systems may well be a shared cognitive infrastructure underlying all modern human languages, both spoken and signed.

Keywords: turn-taking, turn-timing, visual-gestural modality, sign language, sign phonetics, conversation analysis

OPEN ACCESS

Edited by:
Gabriella Vigliocco, University College London, UK

Reviewed by:
Wendy Sandler, University of Haifa, Israel
Simon Garrod, University of Glasgow, UK

***Correspondence:**
Connie de Vos, Language and Cognition Department, Max Planck institute for Psycholinguistics, P.O. Box 310, 6500AH Nijmegen, Netherlands
connie.devos@mpi.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *21 November 2014*
Accepted: *23 February 2015*
Published: *24 March 2015*

Citation:
de Vos C, Torreira F and Levinson SC (2015) Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. Front. Psychol. 6:268. doi: 10.3389/fpsyg.2015.00268

Introduction

Spontaneous conversations among speakers often run smoothly with slight overlaps and gaps between consecutive turns (Sacks et al., 1974). Comparative research has shown that speakers from a broad range of typologically and geographically dispersed languages vary little in response latencies in question–answer sequences, with mean overall offsets at 229 ms, and language-specific means within 250 ms on either side of this cross-language mean (Stivers et al., 2009). A general observation in studies of spoken interaction is that speakers orient toward a one-at-a-time principle when taking turns at talk, and do so at a surprisingly fast pace across a wide range of

spoken languages. The universality of this tightly organized behavior in spoken conversation, as well as its clear precursors in early infancy, make a case for turn-taking constituting an important part of human communicative ethology (Levinson, 2006). A leading question for the research reported here is to what extent sign language users also operate the same turn-taking system as used in spoken languages, especially with regard to turn-timing.

One of the substantial discoveries of the last 50 years is that sign languages show all the properties of full natural languages on all relevant levels of linguistic structure, including, for instance, sublexical structure at the phonological level (Emmorey, 2002; Meier et al., 2002; Sandler and Lillo-Martin, 2006; Channon and van der Hulst, 2011; Berent et al., 2013). However, it has been suggested that, in conversational turn-taking, signers may show greater toleration for overlap due to the characteristics of the visual modality. In signed conversation, visual feedback does not interfere with sign-production, in the same way that auditive feedback does when speaking (Emmorey et al., 2009). This suggests that, if turn-taking was basically motivated by channel limitations (the difficulty of hearing and speaking at the same time), then signed turn-timing should be characterized by a higher proportion of overlapping turns. Additionally, Coates and Sutton-Spence (2001) argue that unlike speakers, signers do not attend to the one-at-a-time principle, and rather form a collaborative floor with their interlocutors, thus having higher degrees of social tolerance for overlap.

Conversation analytic work on sign languages has been sparse, yet early studies of signed conversations have observed some remarkable features of signed interactions. Baker (1977) was the first to note that unlike speakers, sign language users need to ensure visual attention of their interlocutors before they can initiate a turn, hence the use of various types of summonses is more pervasive in signed interaction due to the localized nature of visual information, which requires attentional focus (McIlvenny, 1995). In addition to explicit attention-getting gestures such as waves and taps, this may be done by repeating the initial sign of a turn until recipiency is confirmed through eye contact. The latter strategy appears akin to recycling turn beginnings in spoken conversation, but has the specific purpose of mobilizing recipiency (cf. Schegloff, 1987). In multi-party conversation, multiple signers may self-select for the next turn and thus initiate signing at the same time, although such overlaps may not always be attended to. In addition, there are many other cases of overlapping movement of participants' articulators. As such it is important to consider, in the case of sign language, whether overlapping signs are in fact attended to as competitive turns by speech act participants (McIlvenny, 1995; McCleary and de Arantes Leite, 2013; Groeber and Pochon-Berger, 2014). In the current study, we consider turns that make relevant a timely and contingent response on behalf of the addressee, namely question and answer sequences.

In spoken conversation, addressees sometimes initiate a response at a possible point of completion while the initial speaker continues his turn, thus resulting in *terminal overlap* (Jefferson, 1986; Schegloff, 2000). Similarly, Baker (1977) observed the phenomenon of *partial overlap* at signed turn transitions when "one interactant's hand(s) moving toward the position where a sign will be made as the other interactant is making a sign." However, in considering the comparison between spoken and signed languages, we must bear in mind that the preparation for vocalization in spoken languages is mostly inaccessible to other participants, consisting apart from inbreaths in early motor preparation of the vocal organs (Palo et al., 2014). Therefore, as argued by McCleary and de Arantes Leite (2013), these preparatory movements in sign function on a par with pre-turn inbreaths or other pre-vocal preparation in spoken languages and thus should arguably be excluded from the analysis of the actual exchange of turns. That is to say, it seems reasonable, in order to compare signed, and spoken interactions on equal grounds, to exclude the preparation for signing from timing analysis, as has been the case with the preparation for speaking in the analysis of spoken turn-taking (McCleary and de Arantes Leite, 2013). In our analysis of turn-timing, we differentiate between partial overlaps and possible completion overlaps by taking the start of the initial *stroke* (the 'content' part of the manual gesture) as the turn beginning as it most directly reflects the phonological content of a sign.

Overlap at signed turn transitions may also result from turn-final holds, which are typically released as soon as the relevant response has been recognized (Groeber and Pochon-Berger, 2014). Crucially, signers do not orient to these practices as troublesome in conversation, nor do such overlaps get dealt with using designated overlap resolution devices (cf. Schegloff, 2000; McCleary and de Arantes Leite, 2013). We thus hypothesized that in optimizing turn transitions, sign language users focus on the phonological content of signs as represented by the stroke, and disregard early preparatory movements, and the intentional holding of signs for response, as well as post-utterance retraction. In other words, the end of the final stroke appears to most directly parallel the transition relevance place (TRP) at which a contingent response on behalf of the interlocutor becomes relevant (Sacks et al., 1974).

If turn-taking lies deep in our communicational instincts as has been suggested (Levinson, 2006), then it may be expected to follow broadly similar lines regardless of language modality. We therefore test the prediction that in signed conversations, interlocutors attend to stroke-to-stroke turn boundaries. If this were the case, turn-timing in signed interactions as calculated by stroke-to-stroke turn boundaries should be within the same cross-linguistic range as has previously been reported for spoken languages.

To address this question we analyze turn-timing in 190 question–answer sequences captured from spontaneous conversations of Sign Language of the Netherlands Nederlandse Gebarentaal (NGT). Questions–answer sequences provide a particularly well-suited conversational context in which to investigate turn-timing, as questions make due a conditionally relevant and timely response (cf. Stivers et al., 2009). The signs for each question–answer sequence were coded and checked by native signers for onset, lexical content, and holds and decays using the coding system originally devised for both co-speech gesture and sign language by Kita et al. (1998).

The paper is structured as follows. Section "Materials and Methods" provides details on our data collection as well as

the annotation scheme. Section "Results" presents a statistical comparison of turn-timing in the NGT sample to the spoken languages as reported by Stivers et al. (2009). Finally, section "Discussion" discusses the methodological implications of our work.

Materials and Methods

The NGT Interactive Corpus

This study exploits the NGT Interactive corpus, which consists of spontaneous conversations of native NGT signers in informal settings, which have been collected, and analyzed by Merel van Zuilen, Stephen C. Levinson and Connie de Vos (Max Planck Institute for Psycholinguistics), and Onno Crasborn (Radboud University) from early 2011 onward. All data and analyses have been ethically approved by the Radboud University Ethical committee under the research program *De structuur en ontwikkeling van conversaties in gebarentaal* (De Vos and Levinson; project code ECG2012-1304-098).

The recording sessions of the NGT Interactive corpus took place in participants' homes, at various deaf clubs, as well as a small restaurant, between participants who were long-term acquaintances and friends. These signers were also very familiar with the research assistant who recorded them, and who is a deaf native signer of NGT herself. The data therefore has the character of natural conversation. All conversations were recorded using two HD cameras from different camera angles. The data on which the present study is based features 16 signers (seven females) in one triadic and six dyadic interactions totaling 11 h and 2 min of raw video data. In one of the dyadic interactions, a third person occasionally joined the conversation, but did not participate in any of the question–answer sequences in our analysis. For this reason, this recording is treated as a dyadic interaction. All but one of the signers included in this study had acquired NGT early in life, before the age of ten and all three variants of NGT – Northern, Western, and Southern – are represented in the sample. Conversation topics ranged, unprompted by the investigators, from a work meeting regarding the write-up of a professional paper, home improvement activities, the history of the deaf club, and interpersonal relations.

The video recordings were compressed into MPEG2 format at 1920 × 1080 resolution and 25 fps. The relevant sections were then translated into written Dutch and annotated further using ELAN video annotation software (Crasborn and Sloetjes, 2008). As is customary in sign language research, each sign was glossed using a designated ID-gloss stemming from the Corpus NGT (Crasborn and de Meijer, 2012), and supplemented with novel ID-glosses whenever necessary. Non-manual signals, such as head and body movements, eyebrow movement, and eyegaze were coded in multiple independent tiers.

Identification of Question–Answer Sequences

In order to ensure a diverse sample of question–answer sequences, we selected 30 min segments from each video file that were dense in turn transitions. In identifying these sequences, we adopted the selection criteria which were originally developed as part of the MPI Coding Scheme for Question–Response Sequences in Conversation at the MPI for Psycholinguistics (Enfield et al., 2003) and form the basis of Stivers et al. (2009). NGT polar questions are canonically marked out by raised eyebrows and a head tilt, while content questions are accompanied by a frown (Coerts, 1992). It is also syntactically possible for a content question to be formed in the absence of a wh-sign, as long as the signer uses furrowed brows. More recent work on NGT has also indicated that the brow movements associated with different question types may also be affected by paralinguistic factors, such as affect, and that these non-manual signals are therefore not a reliable cue to syntactic sentence type (De Vos et al., 2009). For these reasons, all questions in our sample were selected based on functional criteria, regardless of whether they made use of an interrogative sentence type. Specifically, we included all turns that evoked an informative answer on behalf of the addressee. Questions that were offered in reported speech, requests for physical actions, rhetorical questions, and two or more questions that were subsequently delivered in a single turn were excluded from the analysis.

Importantly, sign languages are essentially multi-modal in nature in the sense that signers do not only use their hands but also their facial expressions and body postures to express meaning at the linguistic and paralinguistic level. The non-manual components are sometimes considered as the equivalent of intonation or prosody in sign (see for instance van der Kooij et al., 2006; De Vos et al., 2009 on NGT). In the interactions we studied, a facial expression functioned on occasion as a turn on its own, for example when the combined use of a frown with a nose wrinkle and eyegaze at the addressee was taken as an open class repair initiator (similar to spoken *huh?*). Similarly, in some contexts, polar questions evoked a minimal response such as a head nod (*yes*) or a side-to-side headshake (*no*). Stivers et al. (2009) report that in spoken interactions, such visible behaviors result in faster turn transition times compared to vocal-only responses in the majority of languages in their sample. Similarly, non-manual signals in sign languages may often times persist beyond question boundaries and it is unclear at present to what extent each signal should be regarded as part of the turn at talk (De Vos et al., 2009; McCleary and de Arantes Leite, 2013). Our current analyses are therefore focused on the propositional content of the utterance as expressed by the movements of the hands. These manual movements are phonologically specified as part of the language and are most comparable to spoken words as such. Consequently, we have excluded 24 items of the original data set in which either the question or its response were solely expressed non-manually. The remaining set of functional questions were further categorized into polar questions and content questions resulting in a total data set of 190 questions, of which were 104 polar questions and 86 content questions. Overall, polar questions were thus slightly more common than content questions, as is the case in nine out of the 10 spoken language samples analyzed by Stivers et al. (2009:10588). All of the 16 signers that contributed to the corpus are represented in the sample as both questioner and answerer. The triadic conversation included 42 question answer sequences,

whereas the dyadic conversations included 44, 28, 28, 23, 16, and 9 question–answer sequences, respectively.

Movement Phase Coding

Our analysis of turn-timing is based on the coding of the various movement phases that make up a sign. Specifically, we adopted the movement phase coding developed for Sign Language of the Netherlands and co-speech gesture (Kita et al., 1998). This coding system distinguishes four movement phases for each sign: preparation, stroke, hold, and retraction. These movement phases are illustrated in **Figures 1A–D**. This figure displays all four movement phases in relation to the Dutch sign for 'brother,' which is produced by touching the contra-lateral upper arm twice with the middle and index fingers extended. During the preparation phase the hands move into position and the lexically-specified hand shape is selected (**Figure 1A**). The stroke most directly represents the phonological form of a sign and includes the internal movement of a sign, in this case touching the contra-lateral upper arm (**Figure 1B**). Subsequently to the stroke, a signer might hold a sign during interaction, for example to mobilize a response in their interlocutor (**Figure 1C**). Finally, the sign may be retracted and the hands move into resting position (**Figure 1D**).

Importantly, each sign is minimally associated with a single stroke, but the other three movement phases do not always occur (Kita et al., 1998). This analysis is compatible with the view that signs tend to be monosyllabic, consisting typically of a single phase (a path movement and/or a single hand-internal movement) or a repetition of a path movement or hand-internal movement (cf. Coulter, 1982; Brentari, 1998; Sandler, 1999; van der Kooij and Crasborn, 2008). The beginning and end points of each stroke were identified on the basis of the initial and final frame in which the lexically-specified hand shapes for the relevant sign were fully formed. Furthermore, for signs that include a phonologically specified internal movement, the stroke may be lengthened by a repetition of this internal movement, rather than transitioning into an utterance-final hold (cf. Perlmutter, 1992; Nespor and Sandler, 1999; Stewart, 2014). We have also observed this phenomenon our NGT data set, and in such cases only the initial inherent movement, which is lexically-specified, was included into the stroke.

For a subset of items (59 questions) a second coder, who is also a native signer of Sign Language of the Netherlands, applied the same Gesture Phase coding system. Subsequently, any items that showed discrepancies of more than two video frames were discussed, and adjusted when necessary. In a few cases these differences were based in a distinct phonological analysis of the signs that were being used and these two interpretations could not be reconciled. After these discussion sessions the overall correlations between these two coders was 0.9 for the sign-naive boundary measures and 0.98 for the stroke-to-stroke boundary measures.

Phonetic Measures

In our study we report on three phonetic measures of turn transition times based on the coding of gesture phases. The first measure looks at sign-naive turn-boundaries and includes all manual actions, that is, all movement phases that make up a signer's utterance. The second measure looks at stroke-to-stroke turn boundaries, which run from the start of the initial stroke of a turn till the end of the turn-final stroke. For each signer, gestural movement phases from both hands were taken into account. The third and final measure calculates the offset of the addressee's preparation phase with respect to the end of the question's final stroke, and is called signed utterance launch. **Figure 2** illustrates each of the reported phonetic measures schematically. In section "Results," each of these phonetic measures of signed turn-timing are compared to findings from the spoken turn-timing in cross-linguistic study presented by Stivers et al. (2009).

Figure 3 displays an excerpt from the corpus to exemplify the three phonetic measures. This question–answer sequence stems from a conversation between two males who are close friends living in the North of the Netherlands. They are discussing remodeling activities while enjoying lunch on a roof top terrace. Directly preceding the excerpt, the signer to the viewer's right (R) indicates his misunderstanding using a minimal repair initiator by furrowing his brows and leaning forward similar to a spoken '*huh?*' (cf. Dingemanse et al., 2013). As explained above,

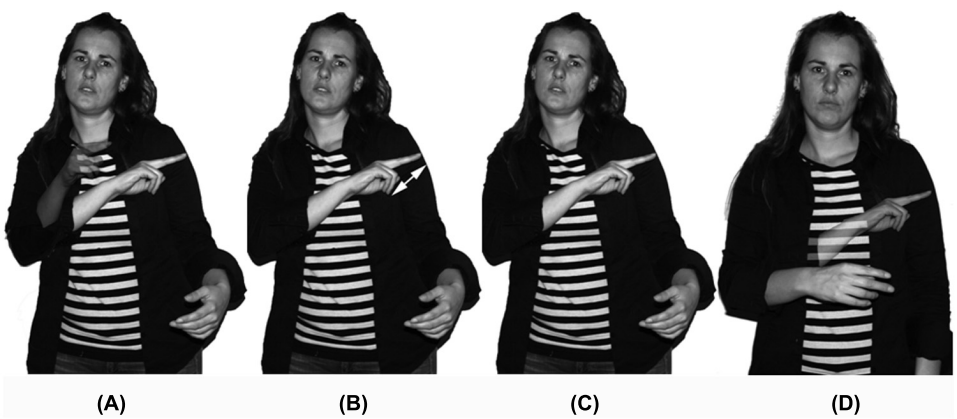


FIGURE 1 | The four gestural movement phases of the Nederlandse Gebarentaal (NGT) sign BROER '*brother*': (A) preparation, (B) stroke, (C) hold, and (D) retraction.

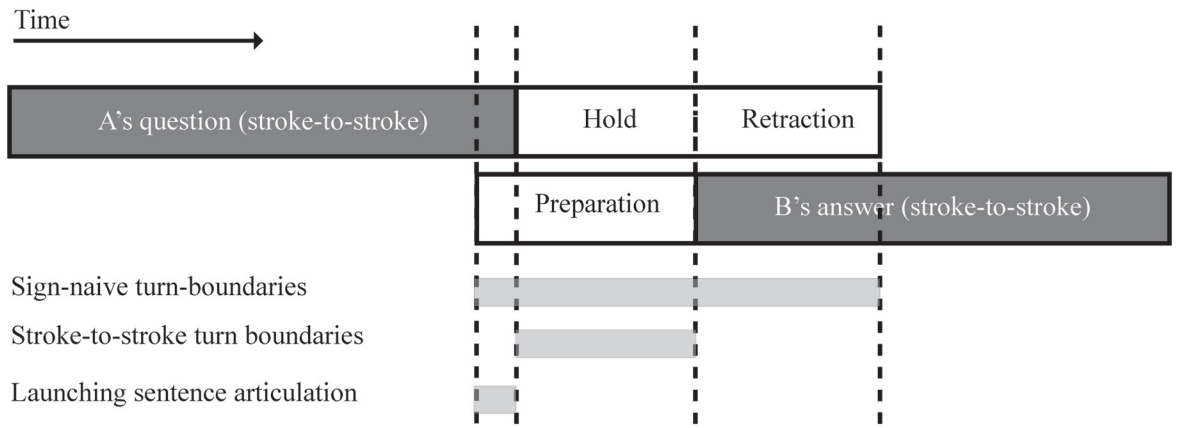


FIGURE 2 | Schematic representation of the three different phonetic measures reported in this study: sign-naive turn-boundaries, stroke-to-stroke turn boundaries, and launching sentence articulation.

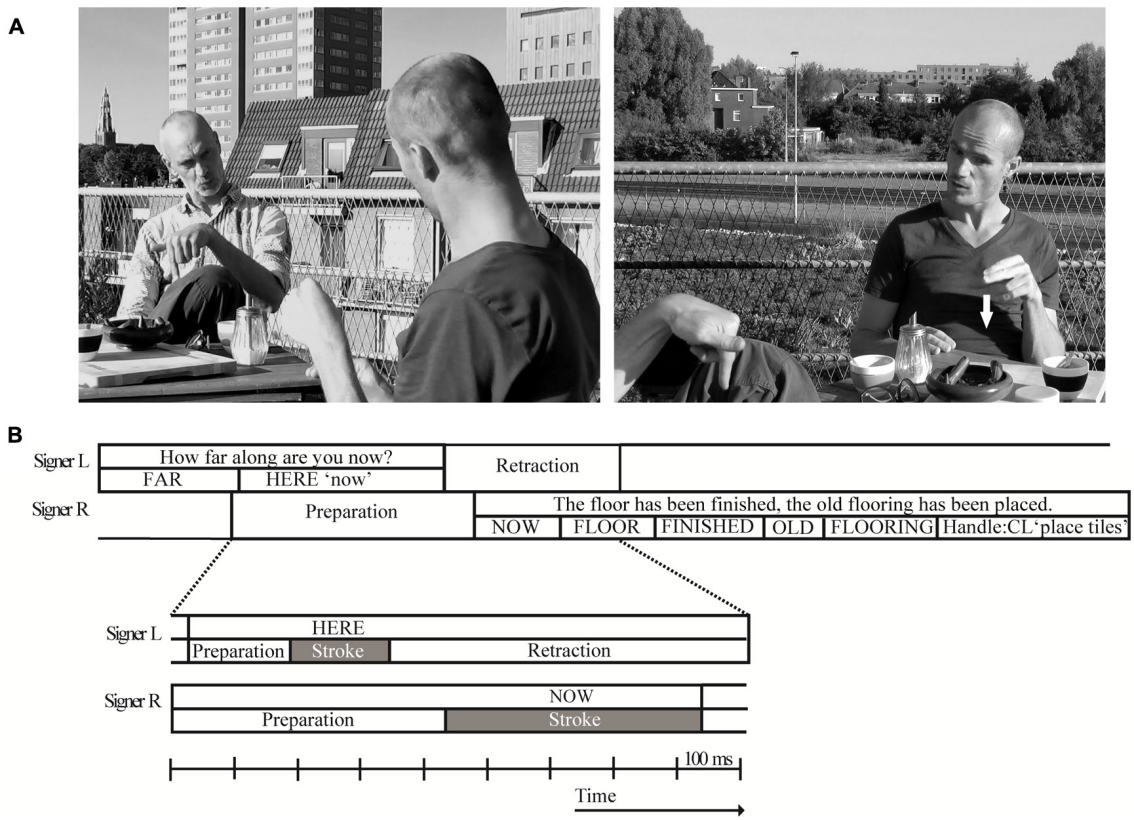


FIGURE 3 | Example of a turn transition from the NGT Interactive corpus: (A) snapshot of a turn transition; (B) scaled representation of the gestural movement phases during the turn transition.

when questions were formed without manual movements these were not included into the current analysis. In this example, each signer uses only a single hand to produce the relevant signs. At the start of the excerpt, the signer on the left (L) asks how far along signer R has progressed regarding the renovation of a particular venue. His question ends in the lexical sign HERE (Dutch: *hier*), which is formed by a downward index finger point, and co-produced with the Dutch mouth movement *nu* 'now.' At the end of his question, the sign HERE is immediately retracted over 430 ms without a sentence-final hold, presumably because the addressee (signer R) has already raised his hand to produce an appropriate response. Signer R's answer starts off with the sign

NOW, and he initiates the preparation of this initial sign before signer L has initiated the start of the preparation of his final sign. As a result, both signers overlap by 690 ms according to the sign-naive turn boundary measure and the launch of R's sentence articulation has a negative value of −260 ms. According to the stroke-to-stroke measurement, however, there is a slight gap of 30 ms between both turns. This turn transition is detailed in (**Figure 3A**) which displays snapshots from both camera angels of the final frame of Signer L's turn-final stroke. On the left-hand side there is a clear view of signer L and the end state of the sign HERE. The camera view on the right-hand side clearly shows signer R who is in the middle of the preparation phase of his turn-initial NOW; the white arrow indicates the trajectory of its stroke. The exact timing of this turn transition is illustrated by a scaled representation of the gestural movements of this turn transition in **Figure 3B**.

Results

The Timing of Question–Answer Sequences

To test whether signers optimize turn-taking on the basis of stroke-to-stroke turn boundaries, we compared turn transition times in NGT with turn transition times in spoken question–answer sequences in ten languages as reported in Stivers et al. (2009). In what follows, we adopt the Floor Transfer Offset (FTO) representation used in De Ruiter et al. (2006), in which gaps are measured in positive milliseconds, overlaps in negative milliseconds. **Figure 4A** shows a density plot of turn transition offset for the overall data set according to the sign-naive turn boundary measure. The sign-naive boundary measure calculates turn-timing by including all gestural movement phases of the hands. According to this phonetic measure, the average offset of answers to questions was −812 ms, the median was −607 ms, and the mode (estimated with the `density()` function in R set to default parameters; R Core Team, 2014, and corresponding to the highest value in the density plot in **Figure 4A**) was −361 ms. These negative values of central tendency indicate that addressees generally start signing well before the question has fully ended. The average value of −687 ms was 6.18 SDs below the cross-linguistic average turn transition time as estimated from Stivers et al. (2009); $m = 229$ ms, SD = 168 ms). Assuming that average turn transition times across spoken languages are normally distributed, and using the data reported in Stivers et al. (2009) to estimate the parameters of this distribution, the probability of observing such a value in this distribution or lower is extremely low ($p < 0.0001$). The sign-naive turn boundary measure thus suggests that the timing of responses to questions in the visual modality deviates substantially from oral-auditory turn-taking in that it exhibits both more and more extended overlap.

Secondly, we calculated turn-timing following the stroke-to-stroke boundary measure, which is based on the hypothesis that signers observe stroke-to-stroke turn boundaries. This phonetic measure calculates the offset of the answerer's initial stroke with respect to the end of the final stroke of the questioner. **Figure 4B** shows a density plot of the stroke-to-stroke turn boundary measure. According to this measure, turn transition times in NGT now exhibit a positive gap, with an average of 307 ms, a median of 269 ms, and an estimated mode of 227 ms. The average value of 307 ms was only 0.46 SD above the cross-linguistic average turn transition time as estimated from Stivers et al. (2009). Assuming that the distribution of average turn transition times across spoken languages is normal, and using the numbers in Stivers et al. (2009), the probability of observing a value of 372 ms or higher is well above the commonly used alpha level of 2.5% for a two-tailed test ($p = 0.32$). When we exclude utterance-initial preparatory movements, and utterance-final holds and decays, turn-timing in signed interaction thus falls within range of oral-auditory turn-taking as reported by Stivers et al. (2009). This is illustrated in **Figure 5**, which shows the average turn transition times in Stivers et al. (2009) plus our two NGT measures.

The Proportion of Overlapping and Delayed Turn Transitions

Another way of looking at signed versus spoken turn-timing is by comparing the distributions of overlapping and delayed turn

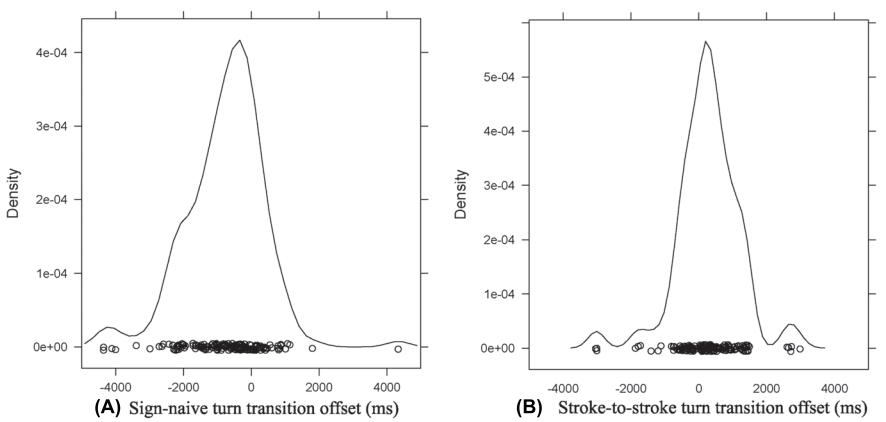


FIGURE 4 | Density plots of answer latency according to the sign-naive (A) and stroke-to-stroke (B) turn boundary measures.

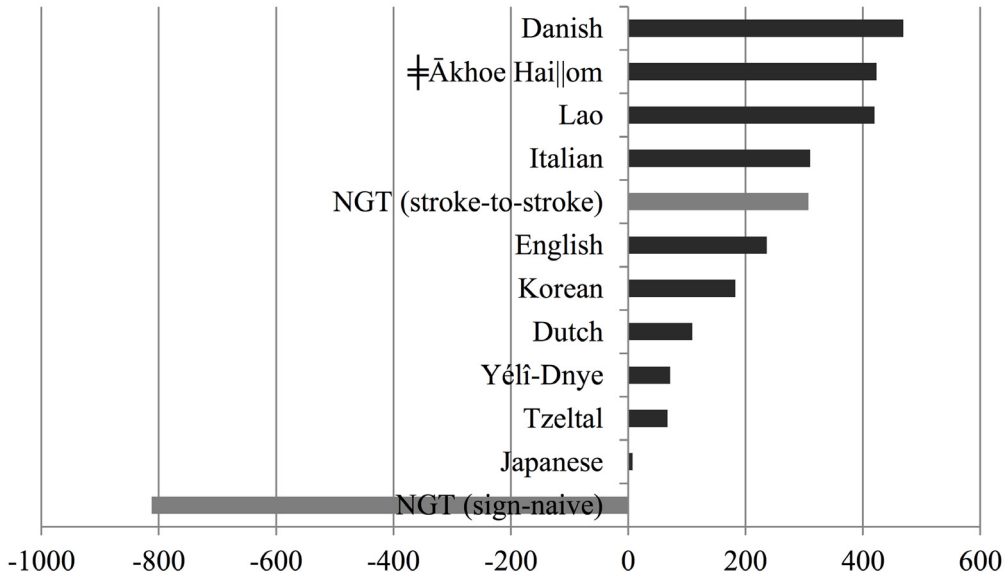


FIGURE 5 | Average turn transition times in 10 spoken languages (in dark gray, from Stivers et al., 2009) and in NGT (in light gray).

transitions. Heldner (2011) previously showed that, with regard to spoken Swedish the threshold for noticeable gaps and overlaps lies at FTOs of 120 ms and −120 ms, respectively. At the moment, it is unknown whether this threshold would generalize to other spoken languages, e.g., to those that have relatively fast or slow turn transition time on average, such as Japanese or Danish, or indeed to sign language. Notwithstanding this caveat, for the sake of comparability, we here consider any turn transition offsets that exceed 120 ms as turns with a noticeable gap, and any turn transition offsets that are −120 ms or less as overlapping in order to compare the distributions of the NGT sample to the Stivers et al. (2009) study.

Figure 6 presents an overview of the percentages of overlapping answers to questions, including the data reported in this study. As estimated by Heldner (2011) based on data from Stivers et al. (2009), the proportion of overlapping turn transitions in question–answer sequences, may range from 13.5% as reported of Lao to 40.0% as reported for spoken Japanese, with a mean of 26.01% and SD of 8.2%. When considering stroke-to-stroke boundaries, 29.8% of responses to questions come in

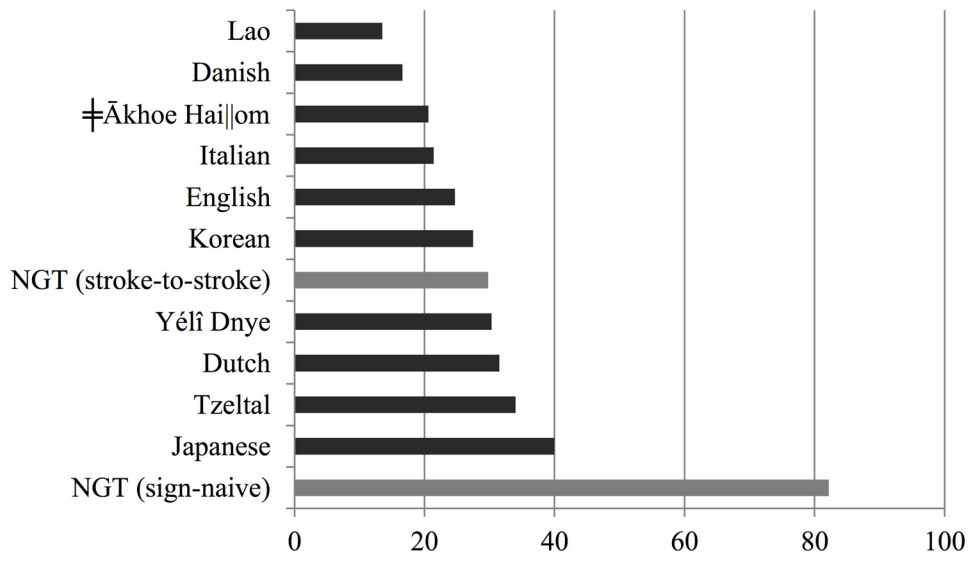


FIGURE 6 | Percentage of turn transitions overlapping 120 ms or more for 10 spoken languages (in dark gray) and in NGT (in light gray). Non-NGT data based on Table II in Heldner (2011:519).

overlap, which is within the cross-linguistic range. Assuming that the distribution of percentages of overlaps in the cross-linguistic sample is normally distributed, and using the data in Stivers et al. (2009) to estimate its parameters, the probability of observing this value or higher in such distribution is well above the commonly used alpha level of 5% ($p = 0.28$). According to the sign-naive measure, however, 82.2% of answers overlap with the respective question. This percentage clearly falls outside the cross-linguistic range for spoken languages. The probability of observing this value or a higher value in the cross-linguistic distribution is extremely low ($p < 0.00001$).

Figure 7 presents an overview of the percentages of transitions between answers and questions involving a gap of more than 120 ms, including the data reported in this study. As estimated by Heldner (2011) based on Stivers et al. (2009), the proportion of delayed turn transitions in question–answer sequences may range from 41.1% as reported of Japanese to 73.0% as reported for spoken Lao. When considering stroke-to-stroke boundaries in the sign language data, 58.3% of responses to questions had a gap of 120 ms or more; this is within the cross-linguistic expectations. Assuming that the distribution of percentages of overlaps in the cross-linguistic sample is normally distributed, the probability of observing a value of 58.3% or lower is well above 5% ($p = 0.54$). According to the sign-naive measure, however, only 17.8% of answers overlap with the respective question, which would be three times less than the spoken language which allows for the smallest number of delays in question–answer sequences. The probability of observing this value or a lower value in the cross-linguistic distribution is much lower ($p < 0.00001$).

Launching Utterance Articulation in Sign

Unlike spoken languages, signed languages offer us the unique opportunity to examine turn preparation using non-invasive methods. In our data, the preparatory phase of utterances had an average duration of 474 ms, a median of 280 ms, and an estimated mode of 255 ms. We computed a third turn transition timing measure corresponding to the latency between the end of the final stroke of the question and the onset of the initial preparatory movements of answer articulation. Relative to the end of the final stroke of the question, response articulation in our data starts with an average latency of −86 ms, a median latency of −78 ms and a modal latency of −53 ms. The timing of preparatory movements in sign is thus slightly earlier than the initiation of pre-utterance inbreaths in answers to questions in spoken Dutch (15 ms; Torreira et al., 2015). We return to this point in the discussion section. **Table 1** summarizes all main results from the present corpus analysis.

Turn-Timing in Dyadic vs. Triadic Interactions

As explained in Section "The NGT Interactive Corpus," one of the recordings in our corpus consisted of a triadic interaction, the other six involving dyadic conversations. Because turn-timing might differ between triadic and dyadic interactions (e.g., due to increased competition for the floor), we examined each of the timing measures presented in the previous subsections (i.e., sign-naive turn boundaries, stroke-to-stroke turn boundaries, and the launching of utterance articulation) as a function of the number of participants in the interaction. We fitted a series of mixed-effects regression models with each of the timing measures as the response, number of participants (dyadic vs. triadic) as a fixed factor, and conversation as a random factor. None of the three models yielded a statistical effect for the fixed factor number of participants ($p > 0.1$ in all three cases), indicating that the triadic conversation was not significantly different from the dyadic conversations in terms of turn-timing.

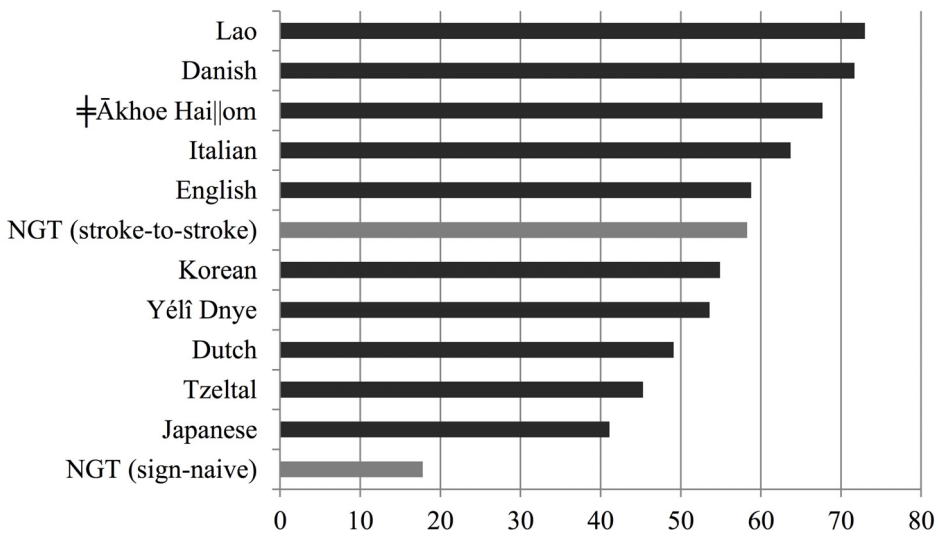


FIGURE 7 | Percentage of turn transitions with a gap of more than 120 ms for 10 spoken languages (in dark gray) and in NGT (in light gray). Non-NGT data based on Table II in Heldner (2011:519).

TABLE 1 | Summary of main results: mean, median, and mode for each phonetic measure of turn-boundaries (in milliseconds).

	Mean	Median	Mode
Sign-naive turn boundaries	−812	−607	−361
Stroke-to-stroke turn boundaries	307	269	227
Launching utterance articulation	−86	−78	−53

Discussion

This study uses a corpus analysis of Sign Language of the Netherlands (NGT) in order to address the question as to what degree turn-timing in signed conversation differs from turn-timing in spoken conversation. The present study has focused on responses to questions, as questions require a timely and contingently relevant response, and could therefore serve as a baseline measure as to how much overlap or gap might be allowed in a given language. Moreover, the timing of responses to questions has been documented in a wide range of spoken languages (Stivers et al., 2009), thus allowing for a controlled comparison to turn-timing across modalities. On the basis of the corpus analysis of spontaneous interactions in NGT, we find that signed conversation exhibits a significantly greater amount of overlap than spoken conversation when we consider all hand action phases as being part of a turn, i.e., preparation, stroke, hold and retraction movements (cf. Kita et al., 1998). Interestingly, however, when we only consider the lexically-specified movement of the hands, i.e., the strokes, turn transition times in signed conversation are clearly within the cross-linguistic range reported for spoken languages, with an estimated mode of 227 ms, and with comparable amounts of gaps and overlaps. We have also found that the timing of the beginning of the preparation phase of the response relative to the end of the last stroke of the question is slightly earlier in signed conversation than that of pre-utterance inbreaths in spoken Dutch conversation, with modes of −53 ms vs. 15 ms.

As noted in the introduction, it has previously been claimed that signed conversation exhibits more overlap than spoken conversation. Coates and Sutton-Spence (2001), for instance, have argued that signers, unlike speakers, may not adhere to a one-at-a-time principle in turn-taking. Rather, they suggest that sign language users are oriented toward a collaborative floor in which more overlap is permitted and socially valued. McCleary and de Arantes Leite (2013) criticized this study for not specifying what types of turns-at-talk may legitimately be used in overlap, and for not providing precise temporal values regarding the use of the various articulators. The present study has shown that, at least in responses to questions, turn-timing in signed conversation looks remarkably similar to that of spoken conversation if we define the delineation of turns on the basis of their stroke phases only (i.e., excluding preparatory, retraction, and hold phases). The decision to consider stroke phases alone is not an arbitrary one, since strokes encode the phonological content of an utterance more directly than other hand movements. In making a comparison between spoken and signed languages, it is therefore plausible that preparatory and retraction movements in signed conversation are best seen as parallel to the pre-beginnings and post-completion elements of spoken turns (cf. Schegloff, 1987), and that TRPs are best approximated by the end of the last stroke. Experimental and qualitative research should address these issues combining the descriptive rigor of Conversation Analysis with perception experiments.

We have also seen that, using the stroke-to-stroke measures, the proportions of turn transitions with noticeable gaps and overlaps in NGT is within the same range as spoken languages according to the stroke-to-stroke boundary measure. These comparisons were carried out on the assumption that the threshold for noticeable gaps and overlaps lies at FTOs of 120 ms and −120 ms, respectively (cf. Heldner, 2011, based on data from Stivers et al., 2009). At present, however, it is unclear as to whether sign languages users are as sensitive to gaps or overlaps as speakers are. These questions could be addressed, for instance, by manipulating turn transition times in pre-recorded signed conversations.

In sign, the articulators are large and heavy, and reaching the articulatory targets of hand strokes from an inactive state will require more time than a vocal articulatory gesture. For this reason, the timing of preparatory hand movements preceding signed turns-at-talk may provide a crucial insight into the time course of signed utterance planning. In our question–answer sequences, we have observed that initial preparatory hand movements of responses are typically launched during the second half of the last stroke of the question, and that the preparatory phase typically ends a couple of 100 ms after the last stroke of the question. On a par with the findings of Torreira et al. (2015) for the timing of pre-utterance inbreaths in spoken conversation, our findings suggest that signers probably attend to final cues to turn-closure when launching their own articulation. If we allow for a reaction time of 200 ms (Fry, 1975), it is plausible that, in the typical case, responders initiate articulation in response to turn-final cues such as final lengthening, which, in sign language, can be manifested as an elongation, repetition, and deceleration of hand movement during the final part of the utterance (cf. Perlmutter, 1992; Nespor and Sandler, 1999; Stewart, 2014). While early cues (e.g., eyebrow movement) in the question may allow for planning the content of the response, local cues close to end of the final stroke (e.g., final lengthening) may provide a general go-ahead signal. The result of this process is a short stroke-to-stroke gap similar to the silent short gaps typically found in spoken question–answer sequences (cf. Stivers et al., 2009), and an overlap interval involving holds, retraction, and preparation phases at turn edges. **Figure 8** shows a schematic representation of typical time courses in a signed question–answer sequence based on modal values in our data (**Figure 8A**), and, for the sake of comparison, and as reported by Torreira et al. (2015) for question–answer sequences in spoken Dutch , in a spoken question–answer sequence in which the answer was produced without a preparatory inbreath (**Figure 8B**), and in a sequence in which a preparatory inbreath was produced (**Figure 8C**). Notice that, in spoken answers too, in line with our findings for sign language, the typical onset of the physical response, in the form of a preparatory inbreath, or of speech proper in answers not

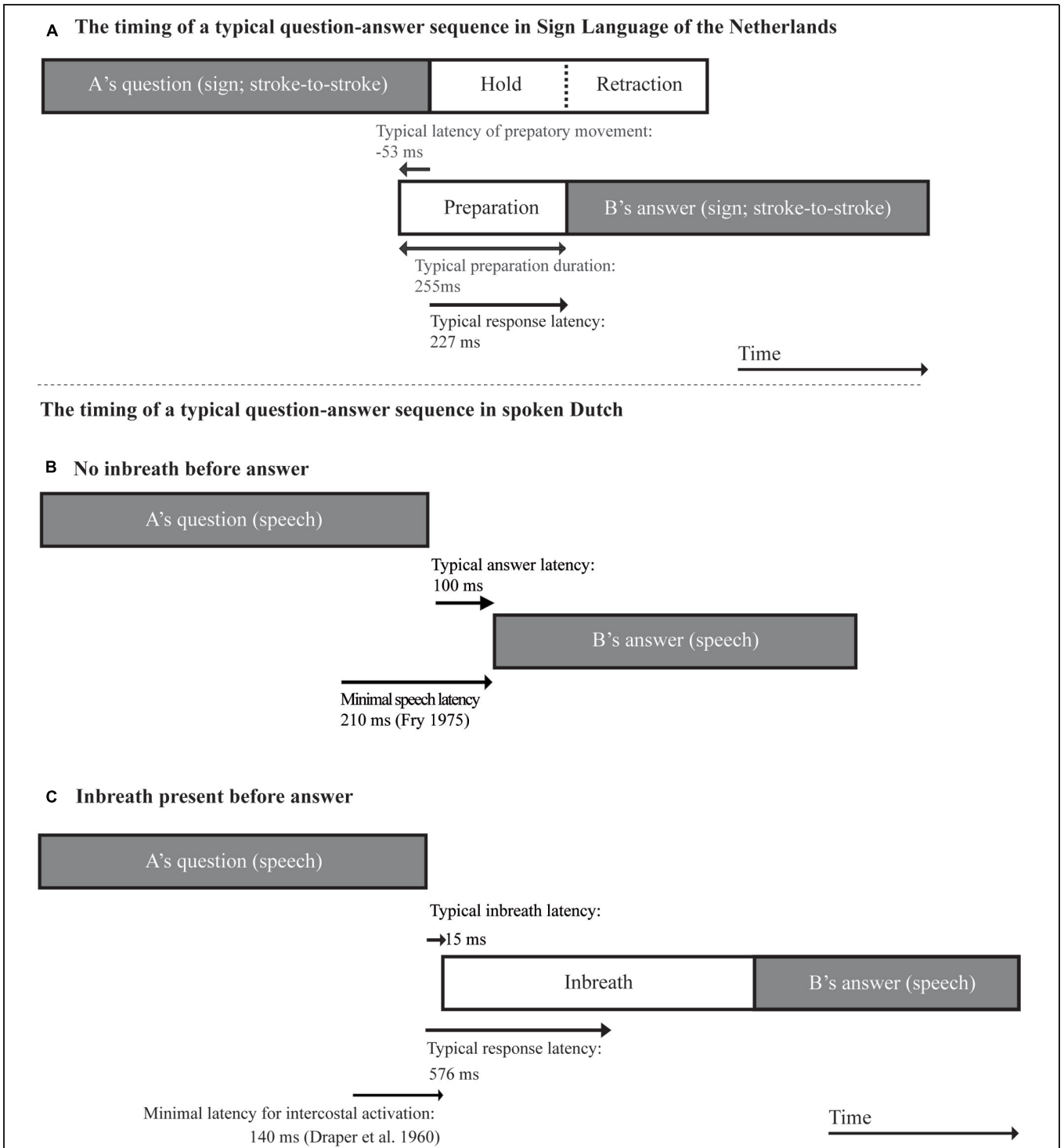


FIGURE 8 | Schematic illustration of typical time courses of: (A) a question–answer sequence in Sign Language of the Netherlands; and a question-answer sequence in spoken Dutch without (B) and with (C) a pre-utterance inbreath (based on Torreira et al., 2015).

preceded by an inbreath, typically starts briefly after the end of the question.

The time course of turn production critically depends on the perception and comprehension of the preceding turn (Levinson, 2013). Recent work suggests that sign-perceivers use early preparatory movements to predict the content of an upcoming sign, resulting in relatively early N400 effects in online signed sentence comprehension (Hosemann et al., 2013). This

phenomenon appears to be afforded by the fact that sign languages exhibit a phonological structure already partly visible in the preparatory movements, thus enabling an early incremental and predictive processing of signs. While pre-utterance inbreaths may indicate the preparation for onset of a spoken turn, pre-utterance signals in spoken language (e.g., lip position) may only offer co-articulatory information about the first segment of the lexical content of an upcoming utterance. It is true that speakers may thus use bodily visible behaviors to enable smooth turn transitions (Mondada, 2007; Ford et al., 2012; Oloff, 2013), but it is unclear at present how routinely these precede vocalization.

A related issue concerns the types of non-manual signals that might enable the accurate projection of TRPs by sign-perceivers. Reaction time experiments are currently being run to establish sign-perceivers' sensitivity to TRPs as defined by stroke-to-stroke turn boundaries (Casillas et al., submitted). That study also aims to identify the visual information signers rely on to determine utterance boundaries online, on the basis of linguistic annotations of the visible cues in this additional data set. On a par with previous work on spoken languages (Ford and Thompson, 1996; Local and Walker, 2012; Torreira et al., 2015), we hypothesize that in addition to lexical content and syntax (De Ruiter et al., 2006; Magyari and De Ruiter, 2012), phonetic and prosodic markers such as signing speed or height (Wilbur, 2009; Russell et al., 2011), as well as visual intonation on the face may play a role (Reilly et al., 1990; Nespor and Sandler, 1999; Fenlon et al., 2007; Dachkovsky and Sandler, 2009; Dachkovsky et al., 2013) in the online prediction of stroke-to-stroke turn boundaries.

This paper has centered on question–answer sequences within a relatively limited data set. Our findings are in accordance with the hypothesis that there may be a single turn-taking system underlying both signed and spoken interactions (Levinson, 2013). If this prediction is borne out by further research, it will extend the discovery that sign languages share all the core features of human language including the domain of communicative turn-taking. Our findings are also consistent with the view that the turn-taking system may be a core part of human communicative ethology, the foundation to language itself. If so, we would expect turn-timing in deaf communities to vary within the range of differences we find across spoken language communities. Of particular interest in this regard are sign languages which have emerged within recent generations such as home sign systems (Goldin-Meadow, 2003), signed contact pidgins (Byun et al., 2014), emerging sign languages (Meir et al., 2010) and rural signing varieties (Zeshan and De Vos, 2012; De Vos and Pfau, 2015). Even though these signing communities have limited time depths, they may follow the same turn-taking principles as other spoken and signed languages. At any rate, we believe that the delineation of turns on the basis of stroke-to-stroke turn boundaries offers a critical tool in the analysis of turn-timing in sign. Preliminary investigations indeed indicate that such an analysis of turn-timing can be instructive to differentiate various types of sequences of consecutive turns to repair misunderstanding in the interactions of signers who do not know a common sign language (Byun et al., 2014; Dingemanse et al., 2014).

In sum, the observed patterns in signed turn-timing in NGT are within the range observed for spoken languages in terms of response latency to questions once we exclude preparatory movements from turn beginnings and retraction phases from turn ends. Moreover, unlike previously argued by Coates and Sutton-Spence (2001), there is now both qualitative (McIlvenny, 1995; Mesch, 2001; McCleary and de Arantes Leite, 2013; Groeber and Pochon-Berger, 2014) and quantitative evidence that sign language users orient to a one-at-a-time principle in taking turns. All in all, our study is consistent with the view that, despite the potential differences between the visual and acoustic language modalities, spoken and signed turn-taking may share more features than has previously been suggested. Further research should center on the question as to what extent the psycholinguistic processes and time course of turn-perception and production in sign versus speech might display similarities as well as differences due to the affordances of each natural language modality.

Acknowledgments

This publication was supported by the Max Planck Gesellschaft as well as the ERC Advanced Grant 269484 INTERACT, awarded to SCL. We would like to thank all signers who contributed to the NGT Interactive Corpus as well as Merel van Zuilen, Johan Ros, and Ellen Nauta who coded the data set, and Mart Lubbers, for his assistance with the processing of the data. We also thank Ellen Nauta for modeling in **Figure 1**, Sean Roberts for comments on the quantitative analysis of the data, and the members of the Language and Cognition Department, especially Elma Hilbrink, Marisa Casillas, and Kobin Kendrick as well as Onno Crasborn and Inge Zwitserlood for comments to previous drafts of this paper.

References

Baker, C. (1977). "Regulators and turn-taking in American Sign Language discourse," in *On The Other Hand: New Perspectives on American Sign Language*, ed. L. A. Friedman (New York: Academic Press), 215–236.

Berent, I., Dupuis, A., and Brentari, D. (2013). Amodal aspects of linguistic design. *PLoS ONE* 8:e60617. doi: 10.1371/journal.pone.0060617

Brentari, D. (1998). *A Prosodic Model of Sign Language Phonology*. London: MIT Press.

Byun, K.-S., Bradford, A., Levinson, S. C., Zeshan, U., and De Vos, C. (2014). Repair sequences in cross-signing: the relationship between try markers and fast track repair sequences. *Talk presented at the International Society for Gesture Studies (ISGS 2014)*, San Diego.

Channon, R., and van der Hulst, H. (eds). (2011). *Formational Units in Sign Languages*. Berlin: de Gruyter and Ishara Press.

Coates, J., and Sutton-Spence, R. (2001). Turn-taking in Deaf conversation. *J. Sociolinguistics* 5, 507–529. doi: 10.1111/1467-9481.00162

Coerts, J. (1992). *Nonmanual Grammatical Markers. An Analysis of Interrogatives, Negations and Topicalisations in Sign Language of the Netherlands*. Netherlands: University of Amsterdam.

Coulter, G. (1982). On the nature if ASL as a monosyllabic language. *Paper presented at the Annual Meeting of the Linguistics Society of America*, San Diego, CA.

Crasborn, O., and de Meijer, A. (2012). "From corpus to lexicon: the creation of ID-glosses for the corpus NGT," in *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, Istanbul, Turkey.

Crasborn, O., and Sloetjes, H. (2008). "Enhanced ELAN functionality for sign language corpora," in *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, Paris.

Dachkovsky, S., and Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Lang. Speech* 52, 287–314. doi: 10.1177/0023830909103175

Dachkovsky, S., Healy, C., and Sandler, W. (2013). Visual intonation in two sign languages. *Phonology* 30, 211–252. doi: 10.1017/S0952675713000122

De Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

De Vos, C., and Pfau, R. (2015). Sign language typology: the contribution of rural sign languages. *Ann. Rev. Linguist.* 1, 265–288. doi: 10.1146/annurev-linguist-030514-124958

De Vos, C., van der Kooij, E., and Crasborn, O. (2009). Mixed signals: combining linguistic and affective functions of eyebrows in questions in sign language of the Netherlands. *Lang. Speech* 52, 315–339. doi: 10.1177/0023830909103177

Dingemanse, M., Blythe, J., and Dirksmeyer, T. (2014). Formats for other-initiation of repair across languages: an exercise in pragmatic typology. *Stud. Lang.* 38, 5–43. doi: 10.1075/sl.38.1.01din

Dingemanse, M., Torreira, F., and Enfield, N. J. (2013). Is Huh? A universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS ONE* 8:e78273. doi: 10.1371/journal.pone.0078273

Draper, M. H., Ladefoged, P., and Whitteridge, D. (1960). Expiratory pressures and air flow during speech. *Br. Med. J.* 1837–1843.

Emmorey, K. (2002). *Language, Cognition, and the Brain: Insights From Sign Language Research*. Mahwah, NJ: Lawrence Erlbaum and Associates.

Emmorey, K., Bosworth, R., and Kraljic, T. (2009). Visual feedback and self-monitoring of sign language. *J. Mem. Lang.* 61, 398–411. doi: 10.1016/j.jml.2009.06.001

Enfield, N. J., de Ruiter, J. P., Levinson, S. C., and Stivers, T. (2003). "Multimodal interaction in your field site: a preliminary investigation," in *Field Research Manual 2003, Part I: Multimodal Interaction, Space, Event Representation*, ed. N. J. Enfield (Nijmegen: Max Planck Institute for Psycholinguistics), 10–16.

Fenlon, J., Denmark, T., Campbell, R., and Woll, B. (2007). Seeing sentence boundaries. *Sign Lang. Linguist.* 10, 177–200. doi: 10.1075/sll.10.2.06fen

Ford, C. E., and Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. *Stud. Interact. Sociolinguist.* 13, 134–184.

Ford, C. E., Thompson, S. A., and Drake, V. (2012). Bodily-visual practices and turn continuation. *Discourse Process.* 49, 192–212. doi: 10.1080/0163853X.2012.654761

Fry, D. B. (1975). Simple reaction-times to speech and non-speech stimuli. *Cortex* 11, 355–360. doi: 10.1016/S0010-9452(75)80027-X

Goldin-Meadow, S. (2003). *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell us About How all Children Learn Language*. New York: Psychology Press.

Groeber, S., and Pochon-Berger, E. (2014). Turns and turn-taking in sign language interaction: the role of frozen signs. *J. Pragmat.* 65, 121–136. doi: 10.1016/j.pragma.2013.08.012

Heldner, M. (2011). Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *J. Acoust. Soc. Am.* 43, 508–513. doi: 10.1121/1.3598457

Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German sign language. *Neuropsychologia* 51, 2224–2237. doi: 10.1016/j.neuropsychologia.2013.07.013

Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Hum. Stud.* 9, 153–183. doi: 10.1007/BF00148125

Kita, S., Van Gijn, I., and van der Hulst, H. (1998). "Movement phases in signs and co-speech gestures, and their transcription by human coders," in *Gesture and Sign Language in Human-Computer Interaction*, eds I. Wachsmuth and M. Fröhlich (Berlin: Springer), 23–35. doi: 10.1007/BFb0052986

Levinson, S. C. (2006). "On the human interaction engine," in *Roots of Human Sociality: Culture, Cognition and Interaction,* eds N. J. Enfield and S. C. Levinson (Oxford: Berg), 39–69.

Levinson, S. C. (2013). "Action formation and ascription," in *The Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Malden, MA: Wiley-Blackwell), 103–130.

Local, J., and Walker, G. (2012). How phonetic features project more talk. *J. Int. Phonet. Assoc.* 42, 255–280. doi: 10.1017/S0025100312000187

Magyari, L., and De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

McCleary, L. E., and de Arantes Leite, T. (2013). Turn-taking in Brazilian sign language: evidence from overlap. *J. Int. Res. Commun. Disord.* 4, 123–154.

McIlvenny, P. (1995). "Seeing conversations: analyzing sign language talk," in *Situated Order: Studies in the Social Organization of Talk and Embodied Activities,* eds P. ten Have and G. Psathas (Washington, DC: International Institute for Ethnomethodology and Conversations Analysis & University Press of America), 129–150.

Meier, R., Cormier, K., and Quinto-Pozos, D. (eds). (2002). *Modality and Structure in Signed and Spoken Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486777

Meir, I., Sandler, W., Padden, C., and Aronoff, M. (2010). "Emerging Sign Languages," in *Oxford Handbook of Deaf Studies, Language, and Education*, Vol. 2, eds M. Marschark and P. Spencer (Oxford: Oxford University Press), 267–280.

Mesch, J. (2001). *Tactile Swedish Sign Language – Turn Taking and Questions in Signed Conversations of Deaf-Blind People*. Hamburg: Signum Verlag.

Mondada, L. (2007). Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Stud.* 9, 194–225. doi: 10.1177/1461445607075346

Nespor, M., and Sandler, W. (1999). Prosody in Israeli sign language. *Lang. Speech* 42, 143–176. doi: 10.1177/00238309990420020201

Oloff, F. (2013). Embodied withdrawal after overlap resolution. *J. Pragmat.* 46, 139–156. doi: 10.1016/j.pragma.2012.07.005

Palo, P., Schaeffler, S., and Scobbie, J. M. (2014). "Pre-speech tongue movements recorded with ultrasound," in *Proceedings of the 10th International Seminar of Speech Production (ISSP 10)*, Cologne, 300–303.

Perlmutter, D. M. (1992). Sonority and syllable structure in American sign language. *Linguist. Inq.* 23, 407–442.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reilly, J. McIntire, M., and Bellugi, U. (1990). The acquisition of conditionals in American sign language: grammaticized facial expressions. *Appl. Psycholinguist.* 11, 369–392. doi: 10.1017/S0142716400009632

Russell, K., Wilkinson, E., and Janzen, T. (2011). ASL sign lowering as undershoot: a corpus study. *Lab. Phonol.* 15, 403–422.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A Simplest systematics for the organization of turn-taking for conversation. *Language (Baltim.)* 50, 696–735. doi: 10.2307/412243

Sandler, W. (1999). Prosody in two natural language modalities. *Lang. Speech* 42, 127–142. doi: 10.1177/00238309990420020101

Sandler, W., and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9781139163910

Schegloff, E. A. (1987). "Recycled turn beginnings: a precise repair mechanism in conversation's turn-taking organization," in *Talk and Social Organisation,* eds G. Button and J. R. E. Lee (Clevedon: Multilingual Review), 70–85.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29, 1–63. doi: 10.1017/S0047404500001019

Stewart, J. (2014). A quantitative analysis of sign lengthening in American sign language. *Sign. Lang. Linguist.* 17, 82–101. doi: 10.1075/sll.17.1.04ste

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Torreira, F., Bögels, S., and Levinson, S. C. (2015). Breathing for answering: The time course of response planning in conversation. *Front. Psychol.* 6:284. doi: 10.3389/fpsyg.2015.00284

van der Kooij, E., and Crasborn, O. (2008). Syllables and the word prosodic system in sign language of the Netherlands. *Lingua* 118, 1307–1327. doi: 10.1016/j.lingua.2007.09.013

van der Kooij, E., Crasborn, O., and Emmerik, W. (2006). Explaining prosodic body leans in sign language of the Netherlands: pragmatics required. *J. Pragmat.* 38, 1598–1614. doi: 10.1016/j.pragma.2005.07.006

Wilbur, R. B. (2009). Effects of varying rate of signing on ASL manual signs and nonmanual markers. *Lang. Speech* 52, 245–285. doi: 10.1177/0023830909103174

Zeshan, U., and De Vos, C. (eds). (2012). *Sign Languages in Village Communities: Anthropological and Linguistic Insights*. Berlin: Mouton de Gruyter. doi: 10.1515/9781614511496

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 18 June 2015
doi: 10.3389/fpsyg.2015.00741

The management of turn transition in signed interaction through the lens of overlaps

Simone Girard-Groeber[1,2]*

[1] Departement Heilpädagogische Lehrberufe, University of Applied Sciences of Special Needs Education, Zurich, Switzerland, [2] Centre de Linguistique Appliquée, University of Neuchâtel, Neuchâtel, Switzerland

There have been relatively few studies on sign language interaction carried out within the framework of conversation analysis (CA). Therefore, questions remain open about how the basic building blocks of social interaction such as turn, turn construction unit (TCU) and turn transition relevance place (TRP) can be understood and analyzed in sign language interaction. Recent studies have shown that signers regularly fine-tune their turn-beginnings to potential completion points of turns (Groeber, 2014; Groeber and Pochon-Berger, 2014; De Vos et al., 2015). Moreover, signers deploy practices for overlap resolution as in spoken interaction (McCleary and Leite, 2013). While these studies have highlighted the signers' orientation to the "one-at-a-time" principle described by Sacks et al. (1974), the present article adds to this line of research by investigating in more detail those sequential environments where overlaps occur. The contribution provides an overview of different types of overlap with a focus of the overlap's onset with regard to a current signer's turn. On the basis of a 33-min video-recording of a multi-party interaction between 4 female signers in Swiss German Sign Language (DSGS), the paper provides evidence for the orderliness of overlapping signing. Furthermore, the contribution demonstrates how participants collaborate in the situated construction of turns as a dynamic and emergent gestalt and how they interactionally achieve turn transition. Thereby the study adds to recent research in spoken and in signed interaction that proposes to rethink turn boundaries and turn transition as flexible and interactionally achieved.

Keywords: conversation analysis, sign language interactions, overlap, simultaneous signing, turn transition, Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*, DSGS)

OPEN ACCESS

Edited by:
Marisa Casillas, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Connie De Vos, Max Planck Institute for Psycholinguistics, Netherlands
Richard Ogden, University of York, UK

***Correspondence:**
Simone Girard-Groeber, Interkantonale Hochschule für Heilpädagogik, Schaffhauserstrasse 239, 8057 Zürich, Switzerland; Centre de Linguistique Appliquée, University of Neuchâtel, Pierre-à-Mazel 7, 2000 Neuchâtel, Switzerland
simone.girard@hfh.ch; simone.girard@unine.ch

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *01 February 2015*
Accepted: *18 May 2015*
Published: *18 June 2015*

Citation:
Girard-Groeber S (2015) The management of turn transition in signed interaction through the lens of overlaps. Front. Psychol. 6:741. doi: 10.3389/fpsyg.2015.00741

Introduction

The precursory work of Sacks et al. (1974) on the machinery of turn-taking in conversation has built the foundation for the conversation analytic tradition. Their ground-breaking paper described how participants in conversation finely coordinate their *turns-at-talk* by minimizing both gaps and overlaps during the transition from a current speaker to a next speaker. Subsequent research in Conversation Analysis (henceforth: CA) has further demonstrated the robustness of Sacks et al.'s (1974) model of turn-taking being achieved on the principle of "one-at-a-time" in interactions involving other languages than English (e.g., Stivers et al., 2009), different contexts (informal and institutional) as well as diverse types of speakers (e.g., L1 and L2 speakers for example by Carroll, 2000; Gardner, 2007).

However, research that pointed to an increased amount of simultaneous talk or of pauses between turn transitions, has also questioned the turn-taking *machinery* as a universal model (as e.g., Tannen, 1984 or Lehtonen and Sajavaara, 1985, cited by Gardner et al., 2009). It was suggested that linguistic and cultural aspects are the reason for such a variation between different turn-taking systems. The present study contributes to this issue by investigating the sequential organization of social interaction in a *signed language*. Signed languages make use of the sequential and simultaneous combination of spatio-visual resources, which can be more or less conventionalized among a specific sign language community. To date there have been relatively few studies on sign language interaction carried out within the framework of CA. Interestingly however, most of the researchers working on signed interaction have to some extent discussed the issue of overlapping signing (e.g., Martinez, 1995; McIlvenny, 1995; Coates and Sutton-Spence, 2001; Lackner, 2009; McCleary and Leite, 2013; De Vos et al., 2015). Some researchers highlight the fact that signed interaction presents a dense occurrence of simultaneous signing, sometimes explicitly questioning the relevance of the one-at-a-time model proposed by Sacks et al. (1974) for signed interaction (Coates and Sutton-Spence, 2001; Lackner, 2009). Other researchers rather emphasize the fact that signers finely coordinate their turn beginnings to potential completion points of current signers (McCleary and Leite, 2013; De Vos et al., 2015).

With this paper I intend to add to this topic by a detailed investigation of those instances in signed interaction where participants actually overlap each other's turns, by focusing specifically on the sequential environment of these overlaps (i.e., onset of the overlap at the beginning, midst or end of the current speaker's turn). This type of investigation sheds light on two issues. First, the analysis of the sequential moment of overlaps aims at revealing whether they present an orderly organization, as it has been shown for overlap in spoken language interaction (cf. Section Overlaps in Spoken and Signed Interaction). Second, the analysis of overlapping signing provides insight into how four participants of a signed language conversation accomplish the actual transitions from one signer to the other. On the basis of a 33-min video-recording of a multi-party interaction with 4 female signers in Swiss German Sign Language (DSGS), the paper shows that (1) signers overlap each other's turns regularly *within possible transition spaces* and not in the midst of syntactic constructions, revealing therefore the same orderliness of overlap as in spoken language interaction; (2) signers actively accomplish smooth transitions between the current and the next signer, collaborating thereby in a situated and collaborative construction of turns. The findings add to recent research in spoken and in signed interaction that proposes to conceive turn boundaries as flexible and interactionally achieved.

I start with providing some details with regard to turn-taking and overlap in signed languages (Section Research on Turn-Taking and Overlap in Sign Language), presenting my conception of turn and further detailing the issue of this study. Then I present the methodology and procedure I followed for the current study (Section Method), specifying the annotation practice and the established categories for analysis. In Section Sequential Environments of Overlapping Signing, I present the results on different types of overlaps before I discuss these findings in Section Discussion.

Research on Turn-Taking and Overlap in Sign Language

The Lexical Unit in Sign Language

The lexical unit in sign language is the manual *sign,* which consists of a distinct combination of four sublexical manual parameters, namely handshape, location (in the signing space), orientation (of the palms) and movement (Boyes Braem, 1990). Non-manual features such as gaze direction or facial expression can have distinctive effect on the lexical level. The sign unfolds sequentially in three basic phases, the *preparation phase* (i.e., hands are brought from rest position to the initial location, orientation and handshape), the *stroke* or *independent hold* (i.e., the proper semantic deployment of the sign) and the *retraction* (i.e., after full deployment the hands are brought back to rest position) (Kita et al., 1998). When annotating signed languages, researchers segment lexical signs in two different ways: either they consider end of one sign to be the start of the next sign (i.e., there is no gap between two signs, the transition from one sign to the other is assigned to the second sign; cf. **Figure 1**), or the start of a sign corresponds to the full deployment of the manual parameters handshape, location and orientation and ends with the end of the stroke, while transition phases are not part of the sign (i.e., there is a gap between two signs; cf. **Figure 2**) (cf. Hanke et al., 2012).

The *Turn* and *Turn Construction Unit* in Conversation Analysis and Sign Language Research

Turn and TCU in Classic CA

In spoken interaction, the beginning and the ending of a participant's contribution represent the delimitation of a *turn*. Each turn can further be built by one or more *turn construction units* (Sacks et al., 1974; henceforth: TCU)[1]. Traditionally, TCUs have been defined as grammatical segments of talk, i.e., lexical, phrasal, clausal or sentential constructions, which are *interactionally relevant*. By interactionally relevant it is meant that participants orient to these units as possibly complete units that end in *transition relevance places* (Sacks et al., 1974; henceforth: TRP), i.e., places where the transition to a next speaker becomes *possible*. This transition is, however, not automatic, because co-participants might not take a next turn and current speakers can extend their turns after the possible completion (with new TCUs or by extending the previous TCU; Sacks et al., 1974; Schegloff, 1996). It is important to note as well that each TCU does not end in a TRP, which is the case in multi-unit turns such as story-tellings. In such cases the TRP can be blocked, for example, by lexical or pragmatic devices (Selting, 2000), e.g., the negotiation of a multi-unit turn through a preface.

[1] As Schegloff puts it, turns are "the organization unit which 'houses' grammatical units" (1996, p. 55).

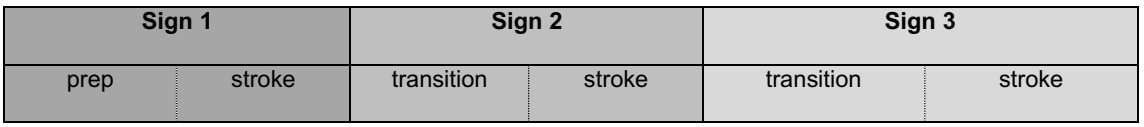


FIGURE 1 | Segmentation of signs including preparation and transition phases.

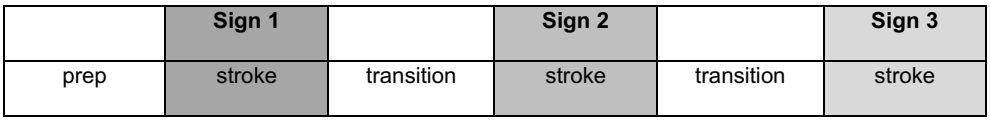


FIGURE 2 | Segmentation of signs excluding preparation and transition phases.

The determination of TCUs has given rise to much debate within the field of CA, about, for example, the relative importance of syntactic and prosodic resources for the indication of completeness of a TCU (e.g., Ford et al., 1996; Schegloff, 1996; Selting, 1996, 2000). The importance of syntactic completion has been demonstrated by Selting (1996). She showed that participants finely time their recipiency tokens or early turn starts to possible syntactic boundaries of a current speaker's turn, showing thereby that they rely on their understanding of syntactic boundaries for the management of turn transitions. A consequence of this is for example that turn-final tag questions (in German "ne"; Selting, 1996, p. 363) are regularly overlapped with next speaker's turn start. However, it is not only syntax that determines whether a TCU may be complete or not. According to Selting, the TCU is a *linguistic unit* constructed with syntactic and prosodic resources, and it is considered linguistically complete in a given sequential and pragmatic context (Selting, 1996, 2000). She states that "*[i]t is the interplay of syntax and prosody that constitutes and delimits TCUs in general*." (2000, p. 489). Put simply, a participant can design a complete syntactic construction but prosodic resources may indicate that there is more to come (cf. rush-through by Schegloff, 1982)—in which case it is not a TCU. In a similar manner, already Local and Kelly (1986) have shown how participants use features such as pitch, loudness and tempo or even glottal stops before a silence for indicating that e.g., a turn is not yet complete (projection of turn continuation) or that an overlapped turn was not complete and will be taken up again after the overlap (Local, 1992).

While TCUs (and turns) are traditionally conceived of as linguistic units, several researchers put into question a logocentric definition of turns-at-talk, where the *vocal production* of *grammatical units* appears to be a basic criterion for being considered part of a turn (cf. recent contributions in Rasmussen et al., 2014). Keevallik (2014) for example proposed an analysis of (non-linguistic) vocalizations as TCUs in their own right (cf. also Goodwin et al., 2000, on the use of nonsense syllables as TCUs or Ford et al., 2012). In a similar vein, an increasing quantity of research on bodily practices shows the relevance of these resources in the construction of "turns" and in the recipient's orientation to these bodily practices as being constitutive for the situated accomplishment of activities (e.g., Goodwin, 2000; Mondada, 2007; Oloff, 2013). In sum, these studies emphasize the fact that social interaction is based on the participants' accomplishment of *actions* by means of ensembles of resources such as speech, gesture, posture and gaze (Kääntä, 2010, proposes to speak of *turns-of-actions*). Within such a perspective, the delimitation of TCUs becomes less important an issue, because the focus of analysis is not on linguistic constructions but on the practices for the organization of activities (Ford et al., 1996).

Turn and TCU in Sign Language Research

Research on signed languages has tried to adopt and adapt the basic notions of conversation analytic research to the situated organization of signed dialogs or social interaction. However, detailed discussions of how the *turn* and the TCU have to be conceived of in signed interaction are still scarce (but see McCleary and Leite, 2013; De Vos et al., 2015). The major challenge for the conception of turns and TCUs in signed languages arises from the fundamental multimodality of signed discourse on the one hand, and the continuous use of lexical, semi-lexical and gestural resources as basic building blocks of meaning construction (Liddell, 2003[2]; König et al., 2012) on the other hand. This resonates very clearly with the current discussions on turn and TCU by those researchers who put into question a logocentric definition of these basic building blocks (cf. Section Turn and TCU in classic CA).

A first fundamental issue in research on signed interaction is how to define the beginning and the end of a turn. Lackner (2009) considered that the lowering and lifting of the forearms constitute turn boundaries. This entails for example that when a participant is holding a sign, he is not yet relinquishing his turn (cf. also Baker, 1977). De Vos et al. (2015) however delineate turns with respect to their stroke phases, excluding therefore the so-called non-verbal movements (holds, preparation phases and retraction phases). This delineation is based on a study that looked at turn transition times with three different phonetic measures (sign-naïve turn boundaries, i.e., transition is measured

[2]Signers make regular use of so-called productive signs, i.e., signs that are only partly conventionalized (they are semi-lexical). They constitute meaning by depicting situations and objects rather than denoting these. Their meanings depend heavily on context.

by considering preparation, stroke, hold and retraction phases; stroke-to-stroke turn boundaries, i.e., the transition is calculated by considering the time between two strokes; and a measure of transition between the next signer's preparation of his first sign in relation to the end of the stroke of the current signer's last sign). They further calculated the latency of these transitions and observed that for the *stroke-to-stroke turn boundary*, transition times were within cultural variation from spoken languages (i.e., 229 ms; Stivers et al., 2009). On the basis of this result they suggest that turns have to be delineated with respect to their strokes, whereas preparation, hold and retraction phases are excluded from the turn. In a study on Brazilian Sign Language, McCleary and Leite (2013) do not explicitly state their delineation of turns. Their analyses show, however, how a current signer self-interrupts his signing in response to gestural movements by an incipient signer, namely a self-groom and a palm-up gesture. Hence, the current signer is clearly orienting to these *non-verbal* resources as constituting overlapping signing that has to be resolved (even if it is not phonological, propositional or lexical). This would be an argument to consider such non-lexical resources as proper parts of signed turns.

In sum, the issue of defining turn boundaries in signed languages traces back to the question of whether non-verbal movements, i.e., preparation phases, holds and retraction phases as well as gestural elements such as palm-ups or even self-grooms should be considered as being part of the turn or not. In the present article turns are delineated in the following way:

(1) The beginning of the *turn* includes the preparation phase of the turn-initial sign. The preparation phase of a turn-initial sign is comparable to a hearable inbreath in spoken interaction, described as a *pre-beginning* element (Schegloff, 1996). Pre-beginning elements are non-lexical elements such as an inbreath, coughs or gestures "*which can serve to initiate a turn, while not yet initiating a TCU within it*." (Schegloff, 1996, p. 93). In a similar vein, Lindström (2006) proposes a differentiation between turn and TCU by considering *presegments* (e.g., pre-beginnings) and post-completions (e.g., *huh*) as not being part of the core TCU, but as a part of the possible turn. Against the background of this, I maintain a differentiation between *turns* as participant's overall contributions including resources beyond grammar, and the grammatical units that such turns (can) host.

(2) The end of the *stroke of a potentially last sign* builds a first TRP, constituting also the end of the core grammatical unit (De Vos et al., 2015). However, I do not consider that the end of the linguistically built unit is automatically the end of the turn. Non-verbal movements that follow this stroke are considered as being part of the *turn*, even if they are not part of the grammatical unit (or the TCU). Consequently, holds after the stroke of the potentially last sign as well as additional gestural elements such as palm-up gestures are considered as being part of the turn. These additional elements extend the turn and finally propose a new TRP (cf. Lindström, 2006 on post-completers).

The Turn-Taking Machinery in Signed Languages

Preliminary and rather detailed observations on turn-taking in sign language can be found in Baker (1977), proposing an account for turn transition between signers based on the sending and interpreting of "signals" for turn regulation. Based on conversations between two dyads of deaf signers in a semi-experimental setting, she provides an overview of *initiation regulators*, *continuation regulators*, and *shift regulators*. The *initiation regulators* comprise those conducts that the incipient signer adopts for displaying that he will initiate a turn. The most fundamental initiation regulator is moving the hands out of *rest position* (in rest position the speaker's hands are relaxed e.g., on table or legs), providing a first visual index that the participant launches a turn. In case of absence of mutual gaze between participants, the incipient signer uses an *attention-getting device* (he waves his hand in the visual field of the addressee, taps on the table or the shoulder of the addressee) in order to establish recipiency. Baker (1977) attaches great importance to the establishment of mutual gaze by stating that the "*speaker cannot initiate a turn until the desired addressee looks at the potential speaker*" (Baker, 1977, p. 221). As already Coates and Sutton-Spence (2001) point out however, it seems that incipient signers regularly launch a turn even without previously established mutual gaze and without explicit attention-getting devices. Resources for this are restarts and holds of sign beginnings, allowing for a smooth turn beginning (cf. Goodwin, 1980 for restarts in spoken interaction). *Continuation regulators* refer to those "signals" that manifest that the signer will continue past a first "*information package*" or after a short pause (Baker, 1977, p. 218). According to Baker, not gazing at the co-participant, speeding up the signing and not returning to rest position as well as holding/freezing the last sign are regulators enabling the current signer to continue. This implies that if a speaker is not willing to abandon his turn, he can either "fill the pause" "*with small movements that indicate planning what next to say or by holding the final position of the last sign*" without gazing at the co-participant (Baker, 1977, p. 227). Finally, *shift regulators* are a set of behaviors used by the current or the incipient signer for signaling that the turn goes over to a next signer. Shift regulators by the incipient signer are understood as signals he deploys in overlap with the current signer's turn. Baker notes that e.g., increased size and quantity of head nods, palm-up gestures or gaze withdrawals are indicating that a co-participant attempts to shift from recipient status to signer status. The current signer's fundamental shift regulator is his gaze toward the potential next signer, as well as a diminution of the sign rate and the return to rest position (see also McIlvenny, 1995; McCleary and Leite, 2013 on overlap). Moreover, according to Baker, the end of a turn is signaled by returning the hands to rest position or by holding the last sign while gazing at the co-participant (cf. also Lackner, 2009)[3].

Overlaps in Spoken and Signed Interaction

Sacks et al. (1974) have shown that participants in social interaction orient to a "one-at-a-time" principle for the

[3]It is however not clear whether she considers the retraction phase and the hold as still being part of the turn or not.

management of turn-taking by minimizing overlapping talk between two or more participants. When overlap was observed it was rather short and regularly occurred at turn-endings. In this environment, Sacks et al. (1974) argued that the overlap manifests the next speaker's endeavor to project "*his start to be earliest possible start at some possible transition-relevance place*" (Sacks et al., 1974, p. 706f). This is particularly relevant in multi-party interaction where other participants possibly compete for the next turn. Somewhat later, Jefferson (1984); Jefferson (1986) undertook detailed analyses of the onset of overlapping turns and the timing between turns. She found that overlap onset regularly occurs within *transition space*[4] (Schegloff et al., 1977), from which she concluded that overlap is in fact an orderly phenomenon. This does not exclude however that overlaps occur also in the midst of turns (not in transition space). However, even in the midst of turns overlapping talk is not necessarily a sign of participants' competition for the floor (French and Local, 1983). Participants can for example exploit simultaneous talk as a resource for creating interactional meaning (e.g., choral co-productions of turns for the accomplishment of agreement or the display of mutual reminiscence, Lerner, 2002). Within word search sequences, co-participants can be provided with the opportunity to overlap so as to re-establish the progression of interaction (conditional access to the turn; Schegloff, 2000). There is also possibility that overlap is actually troublesome, in which case participants repair the simultaneous talk with specific overlap resolution practices (Schegloff, 2000). In all of these cases, the model "one-at-time" is in fact warranted as a basic principle, because deviations from it are either interactionally *meaningful* or *repaired*.

For signed interaction, Cicourel (1973; cited by McIlvenny, 1995, p. 138) suggested that deaf participants may not be "constrained by the sequential ordering or chaining rules, because several signers can allow their signing to overlap continuously and several types of information can be communicated simultaneously which fall under the general notion of kinesic-visual communication." In a similar vein, Baker (1977) stated that the visual mode of interaction in sign language "allow[s] interactants to sign and observe another's signs without a loss of understanding, whereas in oral languages, it is more difficult to hear another's speech while talking." (Baker, 1977, p. 216). In fact, since Baker's seminal paper there have been several discussions especially on the issue of overlap in signed languages. Some scholars clearly contest the validity of Sacks et al.'s "one-at-a-time model" for sign language interactions. On the basis of informal conversations between native signers, Coates and Sutton-Spence (2001) for example observe frequent overlapping and conclude that deaf (female) signers do not orient to the interactional organization of "one speaker at a time" but to a "collaborative floor" as described in Edelsky (1981) for spoken language interaction among women. As previously Coates and Sutton-Spence (2001), also Lackner (2009) observes numerous overlaps in her data of dyadic semi-experimental conversations in Austrian Sign Language. She reports that in 3 of 13 dialogs there is a preferred "successive" structure (i.e., participants orient to the one-at-a-time organization), whereas in 4 of 13 dialogs the participants are constantly in overlap. The rest (6 dialogs) present a varying organization (Lackner, 2009, p. 94). Lackner (2009) also provides some information regarding the characteristics of the overlapping turns (cf. also Martinez, 1995). She mainly differentiates two categories. A first category is built by minimal responses such as GOOD, RIGHT, YES (for the manual minimal responses), which are often realized by smaller movements and in lower sign position than the regular signing space (i.e., they are prosodically attenuated). A second category is formed by the occurrences where the addressee initiates a short turn in overlap with the current signer for the accomplishment of a question, a complement or a comment (Martinez, 1995, p. 94). Lackner mentions that in this case the overlapped signer waits for the end of this overlap and then continues with his turn. Besides these categories of simultaneous signing, Lackner (2009) refers to two other interactional dynamics where a lot of overlaps occur. First, she mentions dialogs where participants accomplish multiple questions or comments during the telling of a participant. She observes that when a current signer is overlapped with a short question, the current signer provides an answer and continues, and the overlapping signer regularly "echoes" this answer so as to display his understanding. This echoing further results in overlap. Second, Lackner (2009) refers to moments when participants "just seem to sign simultaneously," collaboratively constructing the dialog (complementing and referring to each other's turns). According to her, this type of interaction is related to the topic at hand (as e.g., when deafness becomes a topic of interaction). This seems similar to observations of a high-involvement signing style, as made by Coates and Sutton-Spence (2001) on American Sign Language as well as Thibeault (1993, cited by Martinez, 1995) on Filipino Sign Language interaction. All in all, several researchers working on signed interaction point out that overlaps and simultaneous signing are very frequent in signed interaction. They put forward various reasons for this, relating to contextual factors (such as the interactional topic at hand, the high involvement of signers in the interaction and shared experience), physiological factors (overlapping signals do not constrain each other) or cultural factors (women talk, sign language community).

By contrast, McIlvenny (1995) states that sequential organization is also relevant in signed interaction, and that this sequential organization is not so much affected by the fact that sign language is a spatio-visual language. More recently, several studies further pinpointed the signers' orientation to precision-timing and orientation to the coordinated transition between signers on the basis of fine-grained analyses of participants' accomplishment of turn-taking (McCleary and Leite, 2013; Groeber and Pochon-Berger, 2014; De Vos et al., 2015). For example, in a recent conversation-analytic account, McCleary and Leite (2013) provide compelling evidence for several *overlap resolution devices* (Schegloff, 2000) that deaf participants rely upon for managing overlapping signing. These more recent studies have also demonstrated the importance of a clear definition of overlap with respect to the movement

[4]Transition space is "(...) *the environment of a turn's possible completion, at which possible transition to a next speaker becomes relevant*." (Schegloff et al., 1977, p. 366).

phases of signs (cf. Section The Lexical Unit in Sign Language). As already pointed out by McCleary and Leite (2013), in early studies on sign language interaction it is often not clear, whether the overlap between a preparation phase and the stroke of a sign for example has been considered as an overlap or not. Indeed, considering the movement phases of signs/gestures, the following types of overlapping signing can be distinguished.

As it is the stroke or independent hold of a manual sign or gesture that houses the semantic information, an overlap between two strokes may be of a different quality than e.g., an overlap between a stroke and a hold. Indeed, as De Vos et al. (2015) have shown, signers do orient to the end of strokes as turn-boundaries (at least in question-answer sequences), and the overlaying production of retraction, preparation and hold does not seem to be troublesome (cf. also Groeber and Pochon-Berger, 2014). Nevertheless, I suggest that the other simultaneous productions may also fall under the term *overlap* as these movements are also considered as being *part of the turn* (even if they are not part of the syntactic unit; cf. Section Turn and TCU in Sign Language Research). This recalls the difference that is made in spoken interaction between overlapping conducts that are troublesome and others that are not troublesome. For example, in spoken interaction an overlap between the end of a lexical unit by speaker A with the inbreath of the incipient speaker B is of a different quality than the overlap between two lexical units (or overlaps between bodily conducts and grammatical units). An issue on this behalf is that all types of overlaps are indicated with the same transcription symbol (square brackets, [xx])—as mentioned in Groeber and Pochon-Berger (2014), the use of various transcription symbols for different types of overlaps may become relevant for documenting such differences, both in signed and in spoken interaction.

Method

Participants and Data

The current analysis is based on a 33-min four-party interaction in Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*, henceforth: DSGS) between four young deaf women: Denja, Nathalie, Isaline and Melinda. All participants have deaf and signing parents and learned DSGS as their L1. The data comes from a larger corpus of DSGS narratives and interactions that has been gathered within the project "*Gaze and Productive Signing in a Corpus of Interactions of Deaf and Hard-of-Hearing Signers of Swiss German Sign Language (DSGS)*," conducted at the *University of Applied Sciences of Special Needs Education*, and funded by Swiss National Science Foundation. Within this project, participants were invited to attend to a whole day of recording at a film studio producing programs in sign language. While there are some semi-experimental data, the film under study here comes from the lunch break that was recorded with the aim of having a maximally natural interaction. All data was filmed with three cameras. Two cameras were positioned to record two participants frontally, and the third camera captured all participants together.

Procedure and Analysis

Annotation

The annotation was done in *iLex*, a corpus-annotation tool developed at University of Hamburg for sign language documentation (e.g., Hanke, 2002), which the project team uses for the previously cited project. The annotation was then exported in *eaf*-format for an import into the multimedia annotation tool ELAN[5]. For the analysis, several tiers have been added to the ELAN-annotation (cf. Section Analysis).

In the current state, the basic transcript consists of a content translation (done by two interpreters), an annotation of manual signs and participants' gaze conduct. Manual signs have been annotated with *glosses*, i.e., words taken from the spoken language (in our case German spoken language) that roughly describe the meaning of a sign. These glosses function as labels for a sign and are not a precise translation. In the transcript they are always written in capitals. As the project team is working with *iLex*, the glosses in the transcripts are automatically linked to the sign databank for Swiss German Sign Language. When a sign is not available in the lexicon, the annotators create a new gloss, which is automatically added to the lexicon.

In this project we segmented lexical signs in a broad way, i.e., the sign starts with the preparation or the transition phase of the sign, and it ends with the end of the stroke. This implies that the end of sign 1 *is* the beginning of sign 2, and that there are no gaps between signs (cf. **Figure 1**, Section The Lexical Unit in Sign Language; Hanke et al., 2012 on sign segmentation). The annotation by the gloss therefore includes the preparation phase or the transition phase of the sign as well as the stroke of the sign. The stroke starts with the full deployment of the sign's handshape and initial orientation and position. The stroke ends when all manual parameters of the sign are realized (handshape, orientation, location, and movement) (cf. also De Vos et al., 2015). By contrast to De Vos et al. (2015) and McCleary and Leite (2013) we did not include the retraction phase, the return to rest position, into the gloss. For the excerpts under discussion in this article, we added a notation of the gesture phases (Kita et al., 1998) of the manual signs, making the types of overlapping signing more explicit (McCleary and Leite, 2013; De Vos et al., 2015). The detailed transcription conventions are represented in **Table 2**.

A range of other non-manual components such as eyebrows or mouthings[6] are important aspects of sign language. They will be regarded as such in the analysis, but they are not (yet) systematically annotated over the whole corpus.

Analysis

In what follows I briefly outline the different analytical steps that I undertook for the current study once the glosses and gaze conducts have been annotated in ELAN.

[5]The reason for this is that by the moment of analysis, I was more familiar with search functions implemented in ELAN than in *iLex*.

[6]A mouthing is the unvoiced articulation of a word or word part from the spoken language lexicon which is used in the hearing community surrounding the Deaf sign language community (Boyes Braem, 2001, p. 99). In our case, the mouthings are thus German words or word parts.

(1) *Identification of those overlaps relevant for the current study*. As mentioned in Section The Turn-Taking Machinery in Signed Languages, there are different types of overlaps with respect to their quality. In this study the analysis is limited to one specific type of overlap, by excluding:

- When the preparation phase of a participant's sign overlaps with the stroke or the retraction phase of another participant's sign; the reason is that there is no overlap on a semantic level (cf. **Figure 3**, cases b and c in Section Overlaps in Spoken and Signed Interaction; cf. De Vos et al., 2015).
- When a participant is holding a turn-final sign for a moment and another participant produces a sign that is overlapping the hold (cf. **Figure 3**, case d, in Section Overlaps in Spoken and Signed Interaction, cf. Groeber and Pochon-Berger, 2014). The reason is that the end of the stroke may constitute a first TRP, while the hold is extending the turn. There is no overlap on a semantic level.
- When non-manual conducts such as head shake or head nod from one participant overlap the manual sign of another participant. This decision is due to the fact that I did not look at the non-manual turns; whether such types of overlaps should be considered as overlaps to the same extent as overlaps between signs is a question that cannot be addressed in this paper.

Each overlap was tagged in ELAN on a separate analysis tier. On the basis of participants' gaze conduct I then chose to distinguish between two types of overlaps. Most frequently, overlaps occur between participants who are gazing at each other, either at the beginning, the end or over all their turns. This implies that at some point there is mutual orientation and participants can perceive their simultaneous production. Another type of overlap consists of simultaneous signing between two participants while they are addressing a third person through their action and gaze conduct. This type of overlap can go totally unnoticed by the participants (i.e., there is no mutual orientation; cf. also McIlvenny, 1995 on that topic of simultaneous turn beginnings). Such overlaps occur frequently after lapses. The data under study presents a total of 382 overlaps. Out of these, 331 instances occur between two or more participants who are gazing at each other at some point during the overlap. In the remainder of this paper, only those 331 cases are further investigated.

(2) *Description of the overlapping turn with respect to action*. For each overlap I attempted to tag the action that the participants accomplished with that turn. While some turns were easily interpretable on the basis of the *next-turn proof procedure*[7] (Sacks et al., 1974, p. 728), other actions were more difficult to determine, especially in those courses of action that consist of storytellings.

(3) *Description of the sequential environment of the overlapping turn's onset with respect to the overlapped turn*. The analysis of all relevant overlaps resulted in a categorization of the overlaps in three main sequential environments:

[7]The next turn proof procedure is a basic methodological tool within CA for the analysis of interaction within a participant's perspective. For the assignment of actions in interaction, analysts are held to ground their interpretation on how participants themselves understand an action. This necessitates an analysis over three turns, where A is accomplishing an action (e.g., information elicitation), B displays then what he understands this action to be by means of his response (e.g., providing information), and A then either confirms or disconfirms that B has understood adequately.

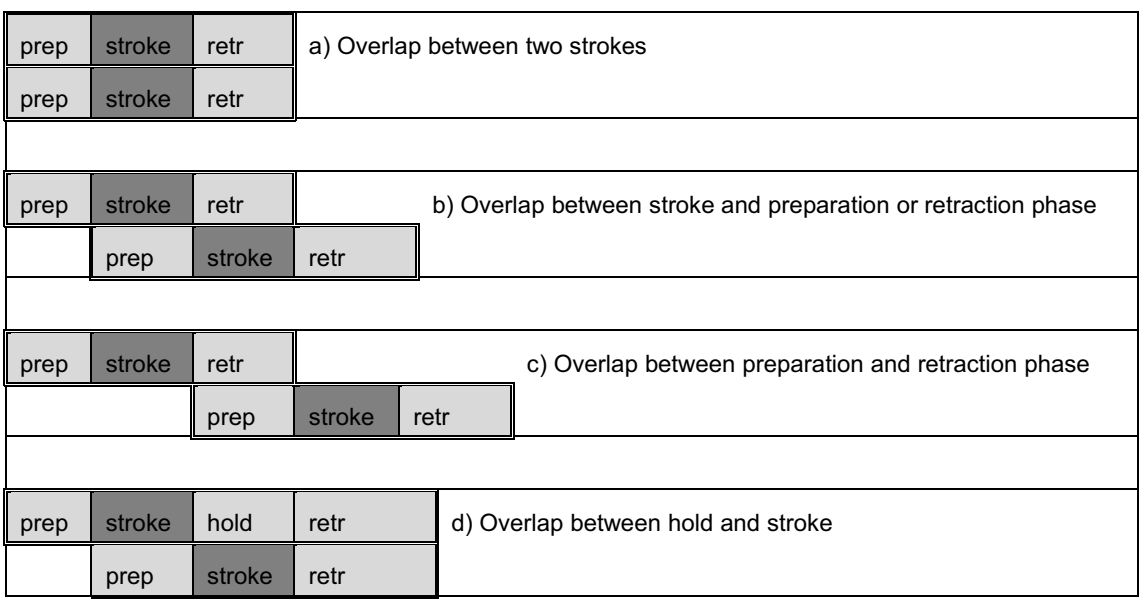


FIGURE 3 | Different types of overlaps.

- The overlap occurs at a first possible completion place reached by participant A (cf. cases A and B in the table below).
- The overlap occurs near the end of a unit or the potential end of a unit (but not exactly at the transition place) (cf. cases C and D in the table below).
- The overlap occurs in the midst of a unit (cf. case E).

Sequential Environments of Overlapping Signing

This section provides an overview of different types of simultaneous signing that can be observed in the data under study. While several authors have commented on the types of actions that are implemented by overlapping turns (e.g., minimal response, repair, short queries, or comments; Martinez, 1995; Coates and Sutton-Spence, 2001; Lackner, 2009), the sequential environments of the overlap onsets have not yet been differentiated systematically in previous studies on signed languages. For this differentiation, I investigate the sequential environments where an incipient signer overlaps an ongoing turn of a current signer. It is important to highlight that the analytic focus is primarily on the sequential environment at the *turn-level*, i.e., whether the overlap occurs at the beginning, in the midst or the end of a possible turn. The overlap onset at the lexical level (i.e., with respect to the movement phases within isolated signs) is visible also in the transcripts, but it is not the focus of analysis in this paper (but see De Vos et al., 2015).

Simultaneous Signing at Places of Possible Completion

One turn environment where simultaneous signing occurs is at first possible completion of a signer's unit (**Table 1**, cases A and B; cf. also Jefferson, 1984). At the precise moment for example where participant A (light gray in **Figure 4**) reaches a possible end of a turn, e.g., the end of the interrogative unit *I HUNGRY* (marked by the end of the stroke of the sign *HUNGRY*), another participant B (gray in **Figure 4**) may launch into a turn. If participant A actually does not continue after this completion, we would observe a smooth turn transition. If however, participant A continues after that first possible completion while participant B also launches into a turn, an overlap occurs[8].

An additional aspect to the overlap onset pertains to the design of overlapped turns. When participant A continues with his turn after a possible completion, this continuation can be independent from the first part (Section Overlapping a Current Signer's New Unit at Places of Possible Completion), or it can be syntactically dependent (Section Overlapping a Current Signer's Dependent Unit at Places of Possible Completion).

Overlapping a Current Signer's New Unit at Places of Possible Completion

The first excerpt illustrates an overlap occurring at a place of possible completion between a current signer, Nathalie, and a new signer, Melinda. The segment stems from a repair sequence initiated by Melinda concerning an exchange that she missed

[8]Note that when A is reaching a possible syntactic completion, it is also possible that, precisely timed to the end of this completion two 'new' participants launch a turn, overlapping each other.

TABLE 1 | Categories of overlap onset tagged in ELAN.

	Category	Overlap onset with respect to the current signer's turn	Explanation (Participant A = current signer; participant B = incipient/overlapping signer)	Quantification
A	Start of new unit	At a first TRP, i.e., after the stroke of the potentially last sign	Both participant A and participant B launch a unit which is syntactically independent from what precedes (new syntactic construction); this can occur after a pause or straightforwardly after a TRP	110/331 (33.2%)
B	Extension of unit	At a first TRP, i.e., after the stroke of the potentially last sign	Participant B launches a turn and participant A adds one or several signs after his initial TRP; these signs are dependent from the first part of the turn; this can occur after a pause or straightforwardly after a TRP	
C	End of unit	During the stroke of the last item of a turn	Participant A produces the last item of his turn (and then retracts his hands), participant B's turn-initial sign (stroke) overlaps the deployment phase (stroke) of that last item	7/331 (2.1%)
D	Potential end of turn followed by continuation	During the stroke of an item that *could* be the last item of the turn	Participant A produces an item that could be the last item of the turn, participant B's turn-initial sign (stroke) overlaps the deployment phase (stroke) of that potentially last item—but after that item A continues with additional signs or with a new (syntactically independent) unit	146/331 (44.1%)
E	Midst of unit	After the beginning of a syntactic unit, not in potential transition space	Participant B launches a turn while participant A's turn is not yet reaching a possible completion	35/331 (10.6%)
F	Undetermined			33 (10%)
			Total overlaps	331

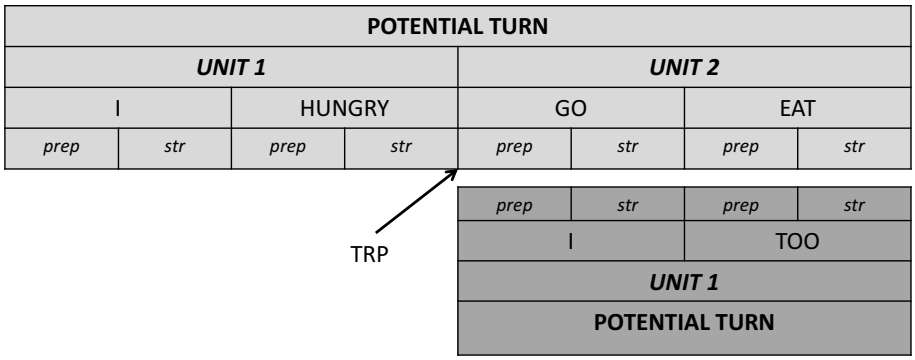


FIGURE 4 | Overlap onset at first possible completion.

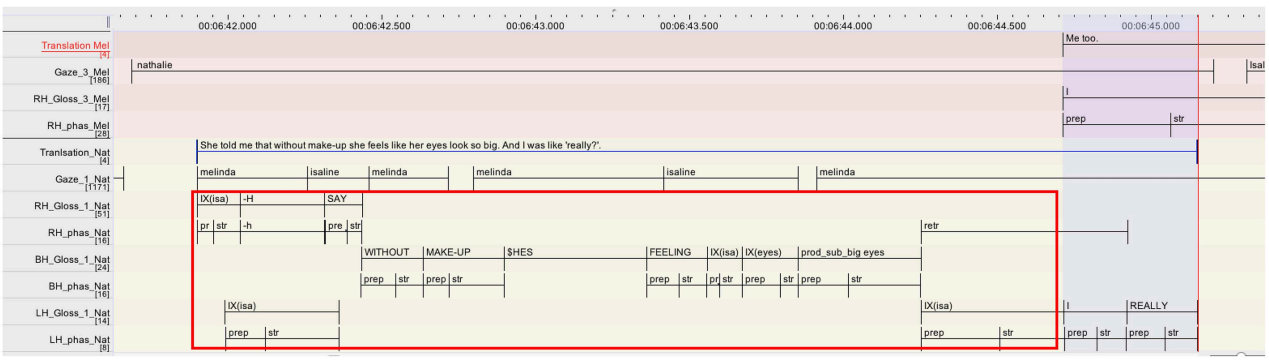


FIGURE 5 | Annotation grid excerpt 1, corresponding to lines 07–09.

between Nathalie and Isaline. Isaline said that she really dislikes her eyes without make-up because they appear so big. Nathalie responds to this by expressing her surprise. During this exchange, Melinda averts her gaze from the participants while she is drinking (she presumably sees however Isaline saying that she dislikes something and the expression of surprise by Nathalie). After Nathalie's display of surprise, Melinda initiates repair by asking Nathalie what Isaline dislikes. When Nathalie reaches a potential completion of her repair, Melinda starts a turn (addressed to Isaline) while Nathalie is continuing with a "new" unit, resulting in simultaneous signing.

The excerpt is presented as follows: a rough gloss annotation of the larger sequence is provided first, followed by a content translation. The part of these transcripts that are in bold are then represented again in the format of the ELAN annotation, so as to provide some more details on the temporal unfolding of the overlap under discussion (**Figure 5**). In the ELAN annotation, each participant is represented with a different color and comprises several tiers. The labeling of the tiers is further explained in the transcription conventions (**Table 2**).

Excerpt 1 (Corpus InterGaze, 00:06:41)

Gloss annotation of larger sequence

01	Isa:	I HATE I NOT MAKE-UP prod-sub-eyes
02		HATE I
03	Nat:	REALLY YOU REALLY
04	Mel:	[WHAT HATE WHAT
05	Nat:	[I LIKE prod-skizz-small eyes LIKE TIRED
06		(FALL-ASLEEP DEAD) I prod-skizz-small eyes WITHOUT
07		**IX(isa) WITHOUT MAKE-UP $HES IX(eyes) prod-sub-big eyes IX(isa)**
08		**[I REALLY**
09	**Mel:**	**[I TOO** I

Translation of larger sequence

01	Isa:	I hate my eyes without make-up – they look so big,
02		I hate that
03	Nat:	Oh really?
04	Mel:	[what does she hate?
05	Nat:	when I don't put make-up my eyes appear very small
06		I look like exhausted without make-up
07		**she told me that without make-up her eyes look so big**
08		**[and I was like 'really?'**
09	**Mel:**	**[me too**

TABLE 2 | Annotation conventions.

TIERS IN ELAN (PER PARTICIPANT)	
Translation_Nat	Tier for content translation of Nathalie's contribution
Gaze_1_Nat	Gaze conduct of Nathalie (participant 1)
RH_Gloss_1_Nat	Gloss for the sign produced on right hand by Nathalie
LH_Gloss_1_Nat	Gloss for the sign produced on left hand by Nathalie
BH_Gloss_1_Nat	Gloss for the sign produced on both hands by Nathalie (for two-handed signs)
RH_phas_Nat	Gesture phases for the sign produced on right hand
LH_phas_Nat	Gesture phases for the sign produced on left hand
BH_phas_Nat	Gesture phases for the sign produced on both hands
TRANSCRIPTION CONVENTIONS FOR THE HAND TIERS RH_GLOSS_1_NATH ETC.	
BALL	Gloss for standardized sign of DSGS
BALL-pl	Plural for BALL
IX(isa)	Pointing toward the person/object in brackets
B-A-L-L	Fingerspelling/Fingerspelled letters (B, A, L)
PALM-UP	A sign/gesture with palms oriented upwards
PALM-DOWN	A sign/gesture with palms oriented downwards
$HES	Hesitation or interrupted sign
D-HAND	Hand configuration of a fingerspelled D, but does not have a clear directional movement as a pointing
prod-man-needle	Productive sign (semi-lexical sign) that consists of a *depiction* of an object, animate referent or a situation (cf. also depicting signs). Productive signs are annotated with the *image producing technique* they use (Langer, 2005) and with a reformulation of what they represent Image producing techniques: **Man**: manipulative technique (cf. also handle classifier) **Skizz**: sketching technique (cf. also size and shape specifier) **Sub**: substitutive technique (cf. also entity classifier)
-H	Hold of a sign/gesture
TRANSCRIPTION CONVENTIONS FOR THE MOVEMENT PHASE TIER	
prep	Preparation of the sign/gesture, i.e., movement out of rest position
str	Stroke
retr	Retraction
-H	Hold
TRANSCRIPTION CONVENTIONS FOR THE GAZE TIER	
D, DD	To the right
G, GG	To the left
D: down	Down
den, nat, isa, mel	Toward Denise, Nathalie, Isaline or Melinda

Nathalie's repair turn (l.07-08) consists of two parts. She starts with reporting that Isaline told her that without make-up Isaline's eyes look so big (red rectangle). This part finishes with the referential pointing toward Isaline (*IX(isa)* on the tier LH_Gloss_1_Nat; l.07) which is doubling the same pointing at the beginning of the unit. This type of reduplication of a pronoun is frequent in several signed languages and has been described as a prosodic marker (e.g., Crasborn et al., 2012). Nathalie then launches into the second unit of her turn with *I REALLY "I said 'really?"* (on the tier LH_Gloss_1_Nat; l.08), reporting her response to Isaline (highlighted in blue).

It is in overlap with this second part of the turn, that the repair initiating party Melinda takes a turn that displays her re-established understanding by affirming that she has the same problem (*I TOO I*, "me too," on the tier RH_Gloss_3_Mel; l.09). Interestingly, Melinda's response to Nathalie's repair is finely tuned to the moment when the element she addressed as repairable has been mentioned, namely *what* does Isaline dislike ("her big eyes"). At this point, she therefore orients to pragmatic completion, as her repair initiation did not ask for anything more than "what does she/Isaline hate." Note also that Melinda launches a response to the repair after the reduplication of the sign *IX(isa)*, where the first part is markedly finished on a syntactic and prosodic level. What we see here is thus precisely what Jefferson (1986) describes as possible completion onset, where "*[a] recipient reasonably, warrantedly treats some current utterance as complete, 'transition ready,' and starts to talk, while (..) the current speaker, perfectly within his rights, keeps going.*" (Jefferson, 1986, p. 154). Considering the movement phases of the overlapping signs, it is interesting to note the following: while Nathalie and Melinda overlap each other by respectively extending and launching a turn, it is only the stroke of *REALLY* (LH_phas_Nat) that is overlapped with a stroke by Melinda (RH_phas_Mel). The other overlaps concern strokes and preparation phases. Interestingly, the preparation phase of Melinda's *I* (RH_phas_Mel) is rather long (350 ms)[9] —it may be possible that Melinda stretches the preparation phase of *I* in response to the fact that Nathalie is extending her turn (overlap resolution device; McCleary and Leite, 2013).

The excerpt illustrates a regular way of turn transition between signers: incipient signers do not necessarily wait for the current signer relinquishing the floor by retracting the hands to rest position. Rather they fine-tune their turn-beginnings to the end of grammatical and prosodic units (cf. also Selting, 1996, 2000 for spoken interaction) marked by the stroke of the turn-final sign. This is similar to the phenomenon observed in spoken interaction, where participants do not normally wait for a pause after a turn for launching a new turn. By contrast, participants finely monitor ongoing turns for their actional, syntactic and prosodic completion[10]. In the present data 33.2% of all overlaps occur at potential completion points where participants reasonably guess that a turn is finished (cf. **Table 1**), while the current participant continues past possible completion.

In excerpt 1, the current signer continues the turn with a second unit that is syntactically independent from what precedes. By contrast, excerpt 2 illustrates a case where the signer's continuation consists of a unit that is dependent on the initial unit.

Overlapping a Current Signer's Dependent Unit at Places of Possible Completion

This type of overlap occurs when participant A reaches potential completion, and participant B launches a turn while participant

[9]Out of 47 annotated preparation phases, 43 instances are less than 250 ms. According to De Vos et al. (2015) this implies that these beginnings are precision-timed even if the preparation phase of participant B is launched after the stroke of participant A.

[10]This is not only the case for transitions between a current and a next signer, but also for cases where a current signer is reaching completion and two new signers launch into a turn.

A continues with one or several further signs that are dependent from the first part. The sequential onset of overlap by the incipient signer is therefore exactly the same as described in Section Overlapping a Current Signer's New Unit at Places of Possible Completion, but the design of the overlapped turn part of the current signer is different.

Excerpt 2 presents a simple case of a short overlap between Melinda and Nathalie within a discussion on the different possible origins of pimples (**Figure 6**). Melinda tells that in the past she had a lot of pimples because of her worries at work (l.04). After this first unit she extends her turn with a second unit, a *PALM-DOWN* gesture, which could be translated as "that's how it was" (l.05). This second part of the turn is overlapped with Nathalie's acknowledging response that can be translated as "you see" with a *PALM-UP* gesture directed toward Melinda (l.06).

Excerpt 2 (Corpus InterGaze, 00:08:12)

Gloss annotation of larger sequenc

01	Nat:	I MAINLY I STRESS I [I D-HAND
02	Mel:	[I TOO I
03	Nat:	[PALM-UP
04	**Mel:**	**[FORMER PIMPLE_pl I COOK WORK**
05		**[PALM-DOWN**
06	**Nat:**	**[PALM-UP(mel)**

Translation of larger sequence

01	Nat:	I have them mainly due to stress
02	Mel:	me too
03	Nat:	[that's life
04	**Mel:**	**[before I had a lot of pimples because of my worries at work,**
05		**[that's how it was**
06	**Nat:**	**[you see**

Nathalie launches her acknowledging response to Melinda's telling, a *PALM-UP* gesture directed to Melinda (RH_Gloss_1_Nat; l.06), in precise overlap with the last item of Melinda's turn, the gesture *PALM-DOWN* (RH_Gloss_3_Mel; l.05). Considering the movement phases of the overlapping gestures, it becomes clear that Nathalie launches her responsive *PALM-UP* near the end of the stroke of Melinda's sign *WORK*, the final sign of her first unit. Thus, Nathalie orients to this moment as reaching potential completion where she can acknowledge Melinda's telling and display her understanding. Melinda's turn however continues with a sign *PALM-DOWN*, which she adds without any manual prosodic disruption (i.e., no pause or slowing down before *PALM-DOWN*). Despite the absence of a manual prosodic break after *WORK*, Melinda deploys a resource that indicates a possible completion after *WORK*—she averts her gaze from her co-participant (to the right side) during the deployment of *WORK* (cf. DD on the tier Gaze_3_Mel). Baker (1977) described gaze aversion from co-participants as recurrent in turn-*beginnings*, where they are exploited as a resource for displaying cognitive planning and holding the turn (cf. also Kendon, 1967). Turn-endings have by contrast have been described as being accompanied with a gaze oriented to the co-participants, indicating thereby that the current participant yields the turn to a next signer (Martinez, 1995; Baker, 1977). This has been observed also as predominant in question-answer sequences in a dyadic teacher-student interaction in Swiss German Sign Language (Groeber, 2011). Both the teacher and the student orient their gaze toward their co-participant at the end of both questions and answers (in 90–100% of cases). An important difference with the excerpt under study here is however that after Melinda's telling (l.04) no projection for a next action is pending (Auer, 2002). The topical talk can continue (as it actually will by Isaline taking a turn), or it could also be closed down. This clearly contrasts with the collection under study in Groeber (2011), where the questions set a strong projection for information provision, while the answers set a projection for an evaluation (cf. three-turn structure in teaching context; Mehan, 1979). In the light of these observations, Melinda's gaze aversion during *WORK* rather indicates the upcoming completion of a sequence similarly to what Rossano (2012) describes for spoken language. Consequently, the addition of *PALM-DOWN* at the end of Melinda's turn can be qualified as a resource that enables Melinda to smoothly step out of her turn. This extension by means of a

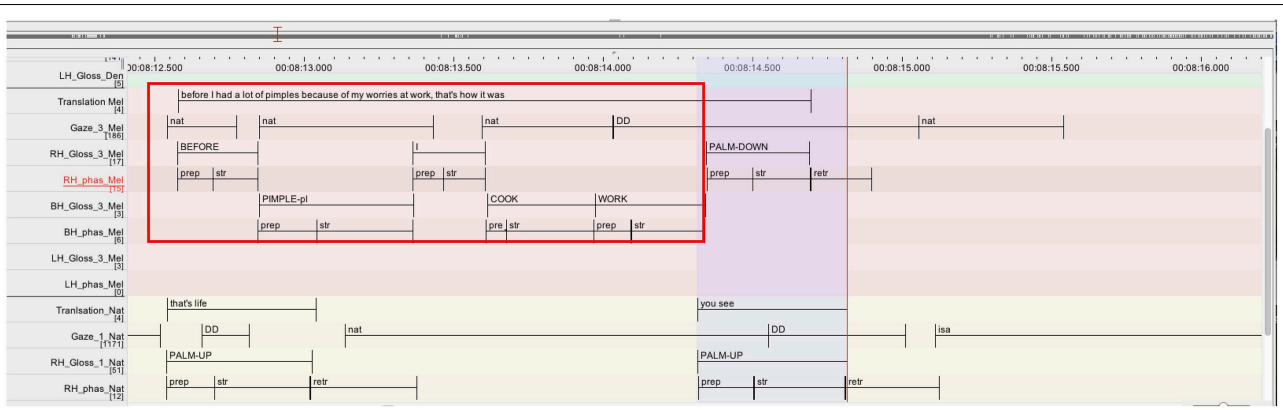


FIGURE 6 | Annotation grid excerpt 2, corresponding to lines 04–06.

short non-lexicalised item results in an overlap that is similar to the turn-final overlaps of tag questions described by Selting (1996; cf. Section Turn and TCU in Classic CA). In fact, the excerpt represents a recurrent way of how signers in the present data end their turns. We will take this issue up again with excerpts 3 and 5.

Excerpts 1 and 2 have shown the fine coordination of incipient signers to possible completion, which implies that the overlapping signing is not the result of mistiming. Signers therefore precisely time their turn beginnings to such places within the temporal unfolding of turns, where syntactic boundaries occur. The overlaps are the result of the fact that the current speaker is not ending her turn at this point, but extends her turn.

Simultaneous Signing after Places of Possible Completion

A further environment where simultaneous signing occurs is just after a possible completion of a signer's unit, i.e., when some pause has occurred after the first possible completion place (i.e., the end of the stroke of the turn-final sign). This pause can consist of a full or partial retraction, or it can be filled with a hold[11]. The case is schematized in **Figure 7** below.

When participant A continues with his turn after some pause after a possible completion, he can do this either by starting a new unit or by adding elements that are somehow dependent on the first part of the turn (as in Section Simultaneous Signing at Places of Possible Completion). In this section I present an example of the first case in excerpt 3 (**Figure 8**). Denja finishes a first unit (l.02), followed by a pause (0.6 s with a hold). After this pause she launches a second unit (l.03) at the same time as her co-participant Nathalie (l.04).

Excerpt 3 (Corpus InterGaze, 00:23.32)

Gloss annotation of the larger sequence:

01	Den:	SECOND SECURITY IMPORTANT GOOD
02		**BUT LEGISLATION D-HAND BAD STOP +(0.6) −H+**

[11] Also for this type of overlap it is possible that two new participants launch a turn after a pause that followed the completion of participant A.

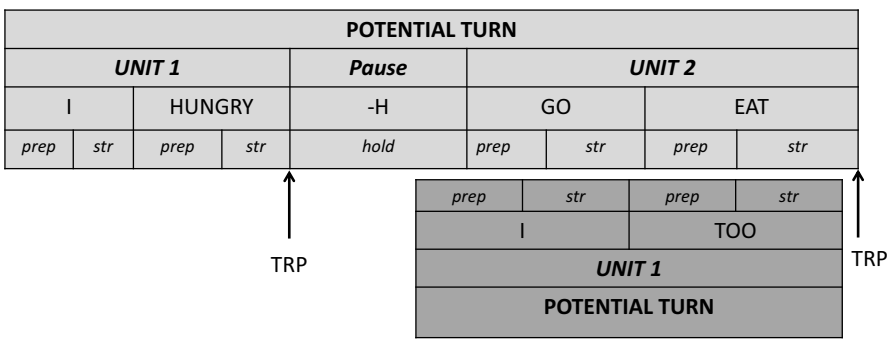


FIGURE 7 | Overlap onset after first possible completion.

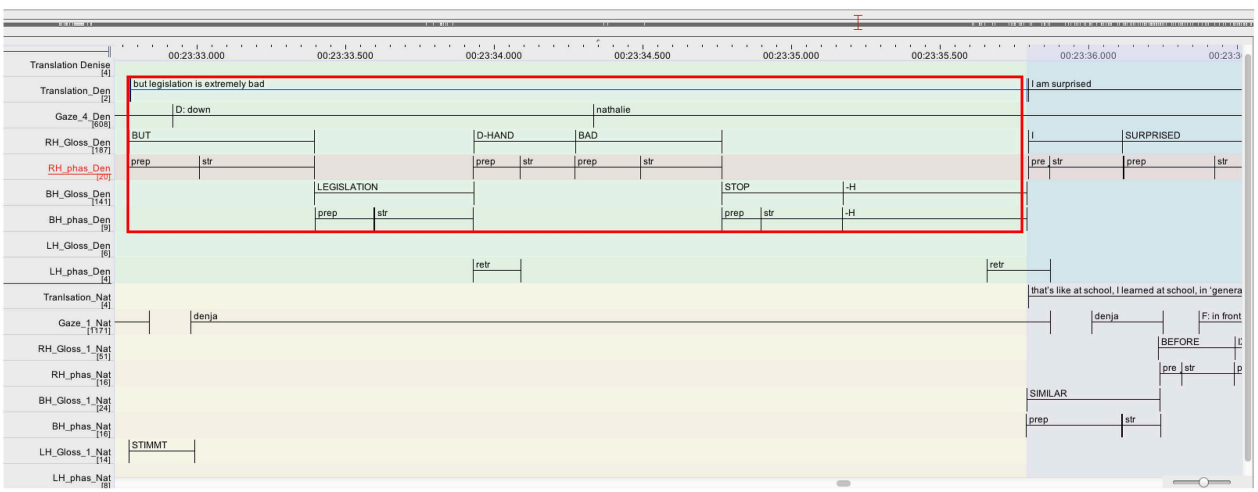


FIGURE 8 | Annotation grid excerpt 3, corresponding to lines 02–04.

03		**[I SURPRISED I +(0.4) −H+]**
04	**Nat:**	**[SIMILAR BEFORE IX(v) SCHOOL I]** SCHOOL I LEARN –A-B-U-[12] I TOO I TOPIC LEGISLATION...

Translation of the larger sequence:

01	Den:	security is important and very good
02		**but legislation is bad**
03		**[I am surprised**
04	**Nat:**	**[that's like at school**
05		**I learned at school,** in 'general knowledge' we also treated the topic legislation ....

The segment starts with Denja affirming that Switzerland is a very safe country, but that there is poor legislation resulting in too mild sentences, as for example for the punishment of rapists or murderers (*BUT LEGISLATION D-HAND BAD STOP -H,* on the tiers RH_Gloss_Den/BH_Gloss_Den; l.01-02, with headshake that starts in parallel to the sign *BAD*). At the end of the sign *BAD*, Denja's turn reaches a first possible syntactic and pragmatic completion. She continues however with the sign *STOP*, after which again speaker change may occur. This sign *STOP* is then held for 0.6 s, and she expands her headshake that she began in parallel to the sign *BAD*). Nathalie provides an acknowledging response by means of a head nod that is precisely timed to the end of the sign *BAD*; this head nod extends simultaneously to the sign *STOP* and the further *-H (hold)*. With this non-manual acknowledgment the course of action has reached some completeness and there is no constraint regarding who of the participants takes a turn. Nathalie might take a full turn during Denja's *-H*, i.e., she could launch a turn going beyond her non-manual acknowledgment, but her hands remain in rest position. By contrast to Nathalie's turn-ending in excerpt 2, where her gaze aversion from the co-participant was interpreted as a possible sequence closing resource, Denja is gazing at Nathalie during her *-H*. I suggest that by means of these resources she invites Nathalie to take a turn and elaborate on the topic (cf. Stivers and Rossano, 2010 for spoken language). In a recent contribution, Groeber and Pochon-Berger (2014) proposed that turn-final holds embody the participant's expectation of the fulfillment of a pending action (as e.g., an answer to a question). In the current excerpt, the turn-final hold also contributes to the embodiment of an expectation for a continuation, even if there is no action projection under way, i.e., there are no constraints (in terms of sequence organization) on how the interaction is expected to continue. Both Nathalie's turn-launching at the end of Denja's hold and Denja's further simultaneous turn extension corroborate this idea. Denja finally continues her turn by adding to her first part of the turn (red rectangle) that this is what she realized (*I SURPRISED I*, tier RH_Gloss_Den; l.03). In overlap with the preparation phase of *I*, also Nathalie finally takes a turn by launching the preparation phase of her turn-initial sign *SIMILAR* (tier BH_Gloss_1_Nat; l.04), orienting thereby to Denja's previously deployed resources (*-H and gaze*) as yielding the floor to her. This simultaneous beginning after the completion point results in overlap (highlighted in blue).

I suggest that from the end of *BAD*, Denja creates a negotiation space (over the signs *STOP* and its *-H*) where participants deal with the determination of a next signer in a subtle and situated manner. This is a most relevant interactional task especially in moments where no projection is pending, which means that there is neither constraint with respect to the next action to be accomplished, nor with respect to whom of the participants will get the floor (cf. also excerpt 2).

It is interesting to note that when Nathalie has finally taken over the turn, Denja does not drop out of the overlap immediately, but she brings her new unit to an end (l.03). However, Nathalie clearly orients to the simultaneous signing as a potentially troublesome overlap, as she restarts the overlapped turn-beginning (*SCHOOL I*, l.04) as soon as the simultaneous signing quits.

In Sections Simultaneous Signing at Places of Possible Completion and Simultaneous Signing after Places of Possible Completion, I have shown that overlaps in signed interaction are, in a lot of cases, orderly (Jefferson, 1984, 1986). They can be a result of the participants' orientation to syntactic and pragmatic completion points of current signers, by launching a turn either at the first possible completion point (stroke of turn-final sign), or slightly past a first possible completion point (after a short pause). In what follows I show that participants can also anticipate an upcoming completion point and launch a turn while the current signer is approaching a first possible completion point (cf. **Table 1**, categories C and D; 153/331 overlaps, 46.2%).

Simultaneous Signing before Places of Possible Completion

The anticipation of an upcoming possible completion can result in different types overlap. Relevant for this study are those overlaps where the stroke of participant A's final sign overlaps with the stroke of participant B's final sign. This is illustrated in **Figure 9** below.

The anticipation can be correct, and the participant A finishes his turn within one sign (as in **Figure 9** and Section Overlapping a Current Signer's Last Item of a Unit), or it can be erroneous and the participant A continues, either with

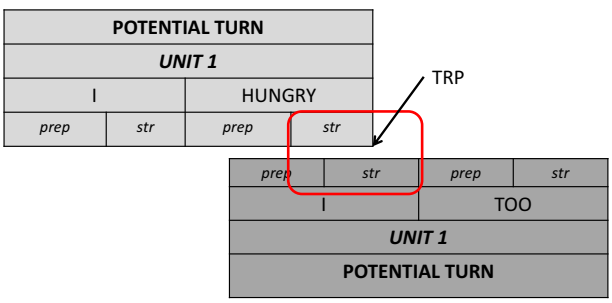


FIGURE 9 | Overlap onset before first possible completion.

[12]ABU is commonly used abbreviation for 'Allgemeinbildender Unterricht' (teaching in general knowledge).

more than one sign, or even with a new syntactic unit (Section Overlapping a Current Signer's Potentially Last Item of a Unit).

Overlapping a Current Signer's Last Item of a Unit

In this section I present an illustration of such cases where a participant B starts signing while participant A is producing the last sign of a unit. These types of overlaps show that participants can foreshadow potential turn-endings. This corresponds to what Jefferson (1986) treats as *terminal onset* and more precisely "*last item*" *onset*, where the final sounds of a last word by a current speaker are overlapped with the beginning of a new speaker. In excerpt 4, Denja and Melinda are talking about Denja's pimple which she has had for 2 weeks (**Figure 10**).

Excerpt 4 (Corpus InterGaze, 00:07:41)

Gloss annotation of the larger sequence:

01	Den:	(XX) THEREUNDER SEE YELLOW THERE IX(pimple) PALM-UP
02		[prod-man-squeeze pimple CAN NOT -H
03	**Nat**	**[(YOU) ALREADY prod-man-needle TRY [YOU**
04	**Isa:**	**[$HES WAVE(isa)**
05	**Den:**	**[ZERO I**

Translation of the larger sequence:

01	Den:	(xx) one sees it's yellow
02		but I cannot squeeze it
03	**Mel:**	**[did you try with a needle [(you)?**
04	**Isa:**	**[euhm denja**
05	**Den:**	**[no I didn't**

Denja is reporting that she was not able to squeeze the pimple despite the fact that it was all yellow. In overlap with this (which is not the focus of this analysis), Melinda is addressing a question to Denja, namely whether she has already tried to pick it with a needle [*(YOU) ALREADY prod-MAN-needle TRY YOU,* tier RH_Gloss_3_Mel; l.03; red rectangle]. Denja launches an answer (*ZERO I*, "no I didn't," tier RH_Gloss_Den; l.05) to this question while Melinda is still producing the stroke of her last sign of her question, the sign *YOU* (l.03; highlighted in blue). Such examples show that signers can project incipient turn completions and launch into the transition even before the current signer has actually finished. Surprisingly, in the data under study here at least, cases where a signer provides minimal responses or launches a new turn in overlap with a sign "under way" (during the stroke deployment) that will actually be the last sign of the turn are very rare (7/331 overlaps; 2.1%; cf. **Table 1**). For those minimal responses and turn starts that arise in overlap, it is much more frequent that they occur during the deployment of a *potentially* last item, i.e., that after that item the current signer continues with her turn (cf. also Section Simultaneous Signing at Places of Possible Completion). In the next section I will focus on those instances in more detail and propose a discussion on the possible reasons for the frequent accomplishment of this type of turn transition.

Overlapping a Current Signer's Potentially Last Item of a Unit

Current signers are recurrently overlapped during a potentially last item of a unit, after which they continue their turn (146/331; 44.1%; cf. **Table 1**). Frequently the overlapping participants are providing only short acknowledgments, hence they are not claiming the floor and the current signer can continue without any disruption. While such turn continuations may have been projected (as e.g., in a storytelling before the climax), other continuations rather seem to occur in the absence of a fuller turn taking by a potential next speaker. Excerpt 5 presents such a case (**Figure 11**). In excerpt 5, Isaline's acknowledging response (l.04) is overlapping the potential end of Nathalie's turn (l.03). The two women are talking about rapists and appropriate prison sentence.

Excerpt 5 (Corpus InterGaze, 00:24:32)

Gloss annotation

01	Nat:	EXAMPLE (NAME) PALM-UP

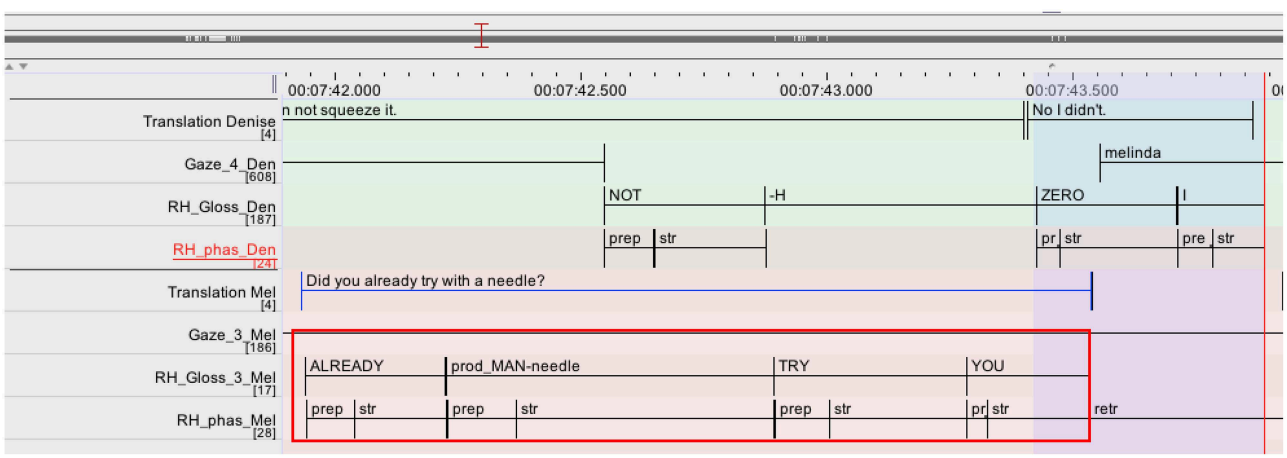


FIGURE 10 | Annotation grid excerpt 4, corresponding to lines 03–05.

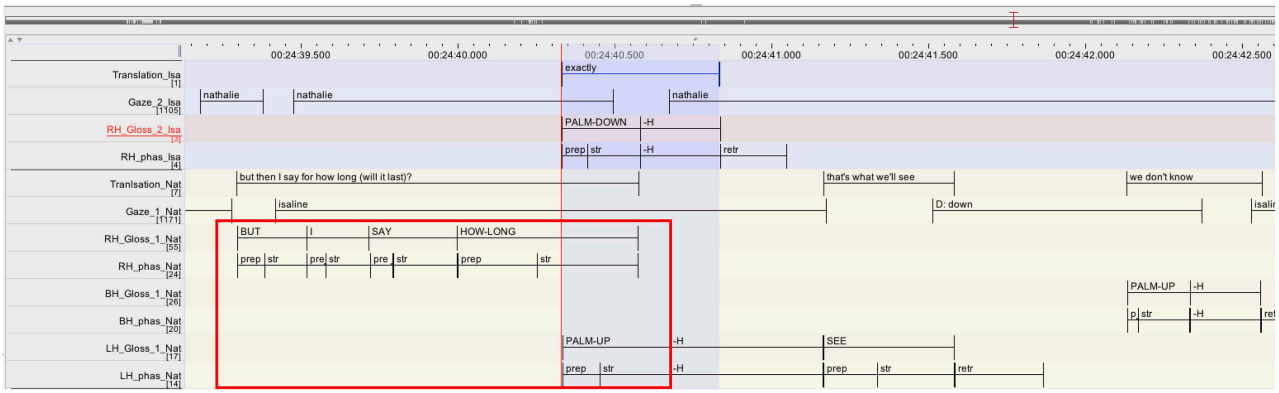


FIGURE 11 | Annotation grid excerpt 5, corresponding to lines 02–06.

02		GOOD PALM-UP
03		**IX(name) BUT I SAY (xx) HOW[-LONG**
04	**Isa:**	**[PALM-DOWN**
05	**Nat:**	**WE-SEE**
06		**PALM-UP**

Translation of larger sequence:

01	Nat:	for example (name)
02		(okay)
03		**(name) but then I say for how long (will it last)?**
04	**Isa:**	**yes exactly**
05	**Nat:**	**that's what we'll see**
06		**we don't know**

Nathalie says that she wonders for how long a rapist who is freed from prison will go before relapse (*BUT I SAY HOW-LONG*, ON TIER RH_GLOSS_1_Nat; l.03). The sign *HOW-LONG* (produced on the right hand) is accompanied with a head tilt back and raised brows (question design; Boyes Braem, 1990). Isaline orients to this moment in Nathalie's turn as potentially reaching some completion; in overlap with the stroke of *HOW-LONG* (l.03) she deploys an acknowledging response that consists of a lifting of a *PALM-DOWN* gesture with a prolonged mouthing of the letter—e-[13] (this could be translated as "exactly" or "yes that's it"). Simultaneously to the preparation phase of Isaline's acknowledgment *PALM-DOWN*, also Nathalie lifts her left hand (while on the right hand she still signs *HOW-LONG*) and produces a *PALM-UP* (tier LH_Gloss_1_Nat; l. 06), which she holds for a moment.

Interestingly, when Nathalie sees Isaline's preparation phase of her acknowledgment (raising the hands out of rest position), Nathalie could relinquish the floor by retracting her hands into rest position (i.e., after the sign *HOW-LONG*). However, Nathalie still prepares and holds her *PALM-UP* on the left hand until Isaline has again retracted her hands to rest position. Different hypotheses can be drawn up for explaining Nathalie's *-H*: first, Nathalie may thereby indicate that she does not want to relinquish the turn, i.e., that she is holding the floor. However, *-H* as a floor-holding device has been described as being accompanied with gaze aversion from the participants (Baker, 1977). Moreover, as the continuation of the excerpt shows, Nathalie will not add a lot of material after this *-H* but rather launch into a closing of the sequence (l.05-06). A second possibility is that, by means of her *-H*, Nathalie embodies her expectation of Isaline's response (Groeber and Pochon-Berger, 2014). However, when Nathalie starts holding her sign, Isaline has already provided a response with her acknowledgment. There is thus not strong evidence for this explanation. A possible explanation that was already mentioned for excerpts 2 and 3, is that Nathalie uses both the *PALM-UP* and the *-H* as resources for a smooth stepping out of the turn, creating thereby the opportunity for a turn transition that is not abrupt but fluid. While Isaline is invited to take a turn beyond her acknowledgment, Nathalie does not immediately end her turn but fills in the delicate moment of transition with these resources. This is further corroborated with Nathalie's continuation in the absence of Isaline's turn-taking. Nathalie extends this negotiation space with the extension by *SEE* ("that's what we'll see," tier LH_Gloss_1_Nat; l.05) during which Nathalie averts her gaze from the co-participant (cf. excerpt 2 for a similar case). It is only after the simultaneous and sequential combination of all these resources that Nathalie retracts her hands to rest position. However, even after having retracted her hands to rest position (on the table), she then further extends the turn with a short *PALM-UP* accompanied with a shrug of the shoulders ("we don't know," tier LH_Gloss_1_Nat; l.06). This is again accompanied with a glance toward Isaline, who still does not take a turn.

In sum, this excerpt showed that Isaline overlaps a possible last item of Nathalie's turn. Nathalie continues with some further signs, which seem to be deployed as a resource for the smooth management of turn transition or turn endings. It is interesting to note that this subtle stepping out of a turn is again occurring

[13]The mouthing –e- may be related to the German word "eben" ("there you have it"), which is also conventionally used with the sign *EBEN* consisting of a palm-up gesture.

at a moment where no specific action projection is set for the continuation of the sequence (cf. excerpt 2).

Up to this point I focused on those overlaps that occur within transition space, corroborating the orderliness of overlapping signing as it has been shown in spoken interaction. A further sequential moment where overlaps occur is what I term *within units*. Jefferson (1986) refers to such overlaps has having an "interjacent onset."

Simultaneous Signing Within Units

A last sequential moment of overlapping signing is when participant B launches a turn in the midst of a syntactic construction by participant A. This type of overlap is rare in the data under study (35/331 overlaps, 10.6%). While the sequential moment of these overlaps does not manifest any coordination between the participants with respect to turn management, an interesting question to ask is what actions participants accomplish by means of these overlapping turns. In what follows I provide a brief description of the principal actional environments in which these overlaps occurred. Note however that each case is very specific with respect to the involvement of the participants and the coordination processes (mutual gaze and orientation).

A first environment where participants overlap each other with interjacent onset is in courses of actions involving repair (7/35 cases). The overlapping party is either self-initiating other-repair, providing a self-initiated other-repair, or displaying his understanding after having launched a repair sequence (sequence-closing third; change-of-state token, Heritage, 1984). As the establishment and maintenance of intersubjectivity is a condition for social interaction to happen, repair appears to be an action that is legitimately accomplished at any place. A second environment of interjacent overlap is built by those moments where participants either display their early understanding (by reformulating a current signer's turn) or by displaying that they share experience with the current signer. By these overlapping turns participants exhibit their alignment with a participant, and at the same time they inform each other about their epistemic status with respect to what is being told (10/35 cases). A third environment is built by courses of action where the overlapping participants display their strong disagreement with a current signer's turn (3/35).

Sometimes participants also provide acknowledgments that are not fine-tuned to transition space but come in midst of a syntactic construction. In 3 of 4 cases, this acknowledgment ensues the introduction of a reference by means of list construction.

Discussion

The Orderliness of Overlaps in Multi-Party Signed Interaction

The turn-taking organization is one of the most basic analytic tools we have in CA for the analysis of broader interactional phenomena (such as repair organization or preference). In order to understand how members of a community accomplish courses of action, and establish and negotiate meaning in a situated and dynamic way, we have to first understand how turns and turn-taking are organized. There is still little research on social interaction in signed languages that adopts a conversation-analytic perspective. To date, we have been provided with some descriptions of how signers indicate their incipient speakership, or their willingness to keep a turn or assign a turn to a next participant. These descriptions do not necessarily follow the analytical mentality of conversation analysis. As a consequence of this, the notions of turn, TCU and TRP are still rarely discussed in detail in this field. Moreover, a most basic issue that has given rise to contradictory assumptions among researchers is the question of whether Sacks et al.'s (1974) turn-taking machinery is also valid for signed languages (e.g., Martinez, 1995; McIlvenny, 1995; Coates and Sutton-Spence, 2001; McCleary and Leite, 2013).

This study did not pursue the question of whether in signed languages there is indeed more overlapping "talk" as compared to spoken interaction. With the analytical focus chosen for this study, the present results provide the field with an initial systematic sketch of the *sequential environments* of overlaps in signed interaction. Using the conversation analytic tools for studying the organization of signed interaction data allowed us to see how participants themselves treat ongoing turns (as being complete or not) and overlaps (as being troublesome or not). The preceding analyses have shown that the vast majority of overlaps produced among four young female signers frequently occur within the sequential environment of possible completion (79.4% of all overlaps; Sections Simultaneous Signing at Places of Possible Completion, Simultaneous Signing after Places of Possible Completion, and Simultaneous Signing before Places of Possible Completion), i.e., they rarely occur in the midst of syntactic units (10.6% of all overlaps; 4.4). This shows that participants finely tune their turn beginnings to those places in the dynamics of interaction, where turns are *possibly complete*. Hence, the majority of overlaps results from the fact that incipient signers anticipate a turn-end and overlap it, and/or that current signers continue beyond a first possible completion. The findings therefore underpin the observed orderliness of overlapping talk in spoken language interaction (Jefferson, 1984, 1986; Schegloff, 2000). Adding to recent findings that demonstrated signers' orientation to precision-timing (De Vos et al., 2015) as well as practices for overlap resolution (McCleary and Leite, 2013), the current study further substantiates the claim that sign language users, too, orient to a turn-taking machinery based on the principle of "one-at-a-time."

The data showed that signed overlap is not messy but organized. Admittedly, however, this does not disprove the assumption that signed interaction presents *more* simultaneous "talk" than spoken interaction, as proposed by Coates and Sutton-Spence (2001) or Lackner (2009). It might well be that signers start their overlaps in an organized manner, and then continue while simultaneously signing for longer stretches than it has been shown for spoken interaction. While I did not systematically investigate the length of the overlaps or overlap resolution practices, the description of what actions participants accomplish with the overlapping turns nevertheless provides us with some valuable insights. Major actions accomplished are acknowledgments, agreements and displays of understanding by means of short reformulations. Furthermore, overlapping turns

were often observed in courses of action involving repair (repair initiations, repairs or change-of-state tokens). Within these actions, overlapping signing does however not imply that the "one-at-a-time" principle is invalid. In fact, participants orient to the fact that the default organization is *one-at-a-time* by keeping their turns short (as in acknowledgments and agreements), or by accomplishing actions that can reasonably overlap ongoing turns because of their urgency (repair initiations) or because of the interactional effects it thereby creates (e.g., strong disagreement). It seems therefore that the observations made on overlaps in spoken interaction are also applicable to the data under study in this article. Consequently, prolonged simultaneous signing and the existence of one-at-a-time principle are not necessarily mutually exclusive.

Gradual Turn Endings and Smooth Turn Transitions

An important finding of the present study is that 'last item overlaps' are scarce (cf. Section Overlapping a Current Signer's Last Item of a Unit). By contrast, participant B often overlaps a potentially last item of participant A who however continues past a possible completion (cf. Section Overlapping a Current Signer's Potentially Last Item of a Unit). This may raise the question of whether participants erroneously predict turn-endings. Another hypothesis would be that current signers who are confronted with overlapping signing by their co-participants continue their turns as a means of holding the floor. While the methodological tools of CA do not allow answering the first question, the sequential analyses of overlapping signing provide some evidence for answering the second one. The analyses have shown that at first places of possible completion, current signers' continuations regularly consist of one or more items, short add-ons. These added items often consist of non-lexical elements or signs that do not contribute substantially new information (*PALM-UP, PALM-DOWN, STOP, -H*). Moreover, after these add-ons, signers often finish their turns. This provides evidence, from a participants' perspective, that there is in fact no attempt to hold the floor. By contrast, on the basis of a fine-grained investigation of the collaborative work accomplished by the signer and his recipient's during possible turn transition, the present analyses suggest that participants deploy such short add-ons as an interactional resource for the management of turn transitions. Concretely, participants step out of their turns in a gradual and smooth manner.

These add-ons can occur after a pause (Section Simultaneous Signing after Places of Possible Completion) or latched to a first possible completion (Section Overlapping a Current Signer's New Unit at Places of Possible Completion; cf. also Section Overlapping a Current Signer's Last Item of a Unit). In the first case, the turn continuation can serve as a resource for treating a problem of recipiency, similarly to turn extensions in spoken language (e.g., Horlacher, 2007). In the second case however, participants may accomplish a slightly different interactional task. When participants continue their turns without any pause after possible completion, they may deal with the fact that neither of their co-participants launches into a turn that goes beyond a minimal response. Hence, the turn extension by a current signer is not oriented to an absence of response, but an absence of a turn-launching that will substantially add to the progression of the activity at hand (cf. e.g., excerpt 6). Concretely, the current signer therefore extends the turn boundary so as to permit a smooth transition without notable pauses between turns.

These observations provide us with some interesting insights regarding turn transition in signed languages. On the one hand, the present study supports that incipient signers orient to the end of strokes/independent holds as first possible transition places where they can launch a new turn (Groeber and Pochon-Berger, 2014; De Vos et al., 2015). On the other hand, as it has been shown for spoken languages, linguistic (syntactic and prosodic) units within a pragmatic context are "possible turns," which can be further expanded with different types of constructions. Thus, turn boundaries are not fixed but flexible. Moreover, and this is a consequence of this first point, the transitions from one signer to the next are also not always clear-cut—within the transition space current and next signers overlap each other as a consequence of the fact that current signers regularly trail-off their turns, stepping out of them in a smooth manner. Consequently, an interesting hypothesis to pursue is the idea that a sense of more overlap in signed interaction may be due to the fact that participants regularly step out of turns in a gradual and smooth manner, rather than ending them with an abrupt retraction of the hands/forearms to rest position. This practice for designing turn transitions may be a specificity of signed languages, but more data has to be investigated to corroborate this idea.

Indeed, it must be emphasized that the present results are limited to one constellation of participants (four acquainted women) and a limited range of courses of action (question-answer sequences, storytellings). Whether the same type of turn transition can be found with other participants (male group, mixed group, unacquainted participants, L2 signers) is an open question. Moreover, as the excerpts in Sections Simultaneous Signing after Places of Possible Completion and Simultaneous Signing before Places of Possible Completion suggest, a systematic analysis of turn-endings that set a strong actional projection as compared to those without strong projection (e.g., questions vs. comments) will be necessary for a more detailed understanding of signers' management of turn-taking.

Acknowledgments

The research leading to these results has received funding from the Swiss National Science Foundation under the grant agreement number 100012_147154/1. I would like to thank all the participants who took part in the project "Gaze and Productive Signing in a Corpus of Interactions of Deaf and Hard-of-Hearing Signers of Swiss German Sign Language (DSGS)." Thanks to Katja Tissi and Sandra Sidler-Miserez for their expert comments on the excerpts. Finally I would like to thank the two reviewers for their detailed and constructive criticism and suggestions on an earlier version of this paper.

References

Auer, P. (2002). Projection in interaction and projection in grammar. *Interact. Linguist. Struct.* 33, 1–43.

Baker, C. (1977). "Regulators and turn-taking in American sign language discourse," in *On the Other Hand*, ed L. Friedman (New York, NY: Academic Press), 215–236.

Boyes Braem, P. (1990). *Einführung in die Gebärdensprache und Ihre Erforschung.* Hamburg: Signum Verlag.

Boyes Braem, P. (2001). "Functions of mouthings in the signing of deaf early and late learners of Swiss German Sign Language (DSGS)," in *The Hands are the Head of the Mouth,* eds P. Boyes Braem and R. Sutton-Spence (Hamburg: Signum), 99–131.

Carroll, D. (2000). Precision Timing in Novice-to-Novice L2 Conversations. *Issues Appl. Linguist.* 11, 67–110.

Cicourel, A. (1973). *Cognitive Sociology: Language and Meaning in Social Interaction.* London: Macmillan.

Coates, J., and Sutton-Spence, R. (2001). Turn-taking patterns in Deaf conversation. *J. Sociolinguist.* 5, 507–529. doi: 10.1111/1467-9481.00162

Crasborn, O., Van der Kooij, E., and Ros, J. (2012). On the weight of phrase-final prosodic words in a sign language. *Sign Lang. Linguist.* 15, 11–38. doi: 10.1075/sll.15.1.02cra

De Vos, C., Torreira, F., and Levinson, S. (2015). Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. *Front. Psychol.* 6:268. doi: 10.3389/fpsyg.2015.00268

Edelsky, C. (1981). Who's got the floor? *Lang. Soc.* 10, 383–421.

Ford, C. E., Fox, B. A., and Thompson, S. A. (1996). Practices in the construction of turns: the 'TCU' revisited. *Pragmatics* 6, 427–454. doi: 10.1075/prag.6.3.07for

Ford, C. E., Thompson, S. A., and Drake, V. (2012). Bodily-visual practices and turn continuation. *Dis. Process.* 49, 192–212. doi: 10.1080/0163853X.2012.654761

French, P., and Local, J. (1983). Turn-competitive incomings. *J. Pragmat.* 7, 17–38. doi: 10.1016/0378-2166(83)90147-9

Gardner, R. (2007). "'Broken starts': bricolage in turn starts in second language talk," in *Language Learning and Teaching as Social Interaction*, eds Z. Hua, P. Seedhouse, L. R. Wei, and V. Cook (New York, NY: Palgrave Macmillan), 58–71.

Gardner, R., Fitzgerald, R., and Mushin, I. (2009). The underlying orderliness in turn-taking. Examples from Australian talk. *Aust. J. Commun.* 36, 66–89.

Goodwin, C., Goodwin, M. H., and Olsher, D. (2000). "Producing sense with nonsense syllables: turn and sequence in the conversations with a man with severe aphasia," in *The Language of Turn and Sequence*, eds B. Fox, C. Ford, and S. Thompson (Oxford: Oxford University Press), 56–80.

Goodwin, C. (1980). Restarts, pauses, and the achievement of mutual gaze at turn-beginning. *Soc. Inq.* 50, 272–302. doi: 10.1111/j.1475-682X.1980.tb00023.x

Goodwin, C. (2000). Action and embodiment within situated human interaction. *J. Pragmat.* 32, 1489–1522. doi: 10.1016/S0378-2166(99)00096-X

Groeber, S. (2011). "Understanding social interaction with bimodal bilingual language learners: focus on turn-taking," in *Talk at the 33. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft* (Göttingen).

Groeber, S. (2014). *Interactional Competence of Hard-of-Hearing Adolescents: The Management of Progressivity in an Integration Classroom.* Unpublished Doctoral Thesis, University of Neuchâtel.

Groeber, S., and Pochon-Berger, E. (2014). Turns and turn-taking in sign language interaction: a study of turn-final holds. *J. Pragmat.* 65, 121–136. doi: 10.1016/j.pragma.2013.08.012

Hanke, T. (2002). "iLex - A tool for sign language lexicography and corpus analysis," in *Proceedings of the 3rd International Conference on Language Resources and Evaluation* (Paris: ELRA), 923–926.

Hanke, T., Matthes, S., Regen, A., and Worseck, S. (2012). "Where does a sign start and end? Segmentation of continuous signing," in *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon Language Resources and Evaluation Conference (LREC)* (Istanbul), S.69–74.

Heritage, J. (1984). "A change-of-state token and aspects of its sequential placement," in *Structures of Social Action*, eds J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 299–345.

Horlacher, A.-S. (2007). La dislocation à droite comme ressource pour l'alternance des tours de parole: vers une syntaxe incrémentale. *TRANEL* 47, 117–136.

Jefferson, G. (1984). "Notes on some orderlinesses of overlap onset," in *Discourse Analysis and Natural Rhetoric*, eds V. D'Urso and P. Leonardi (Padua: Cleup Editore), 11–38.

Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Hum. Stud.* 9, 153–183. doi: 10.1007/BF00148125

Kääntä, L. (2010). *Teacher Turn-Allocation and Repair Practices in Classroom Interaction. A Multisemiotic Perspective.* Ph.D. Thesis. University of Jyväskylä.

Keevallik, L. (2014). Turn organization and bodily-vocal demonstrations. *J. Pragmat.* 65, 103–120. doi: 10.1016/j.pragma.2014.01.008

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychol.* 26, 22–63. doi: 10.1016/0001-6918(67)90005-4

Kita, S., Van Gijn, I., and Van der Hulst, H. (1998). "Movement phases in signs and co-speech gestures, and their transcription by human coders," in *Gesture and Sign Language in Human-Computer Interaction*, eds I. Wachsmuth and M. Fröhlich (Berlin: Springer), 23–35.

König, S., Konrad, R., and Langer, G. (2012). "Der Wortschatz der DGS," in *Handbuch Deutsche Gebärdensprache*, eds H. Eichmann, M. Hansen, and J. Hessmann (Hamburg: Signum), 111–164.

Lackner, A. (2009). *Turn-Taking und Dialogstruktur in der Oesterreichischen Gebärdensprache. Eine Gesprächsanalyse der Salzburger Variante*, Vol. 81. Hamburg: Das Zeichen.

Langer, G. (2005). *Bilderzeugungstechniken in der Deutschen Gebärdensprache* (Hamburg: Das Zeichen), 254–270.

Lehtonen, J., and Sajavaara, K. (1985). "The silent Finn," in *Perspectives on Silence*, eds D. Tannen and M. Saville-Troike (Norwood, NJ: Ablex Publishing Corporation), 193–201.

Lerner, G. (2002). "Turn-sharing. The choral co-production of talk-in-interaction," in *The Language of Turn and Sequence*, eds C. E. Ford, B. A. Fox, and S. A. Thompson (New York, NY: Oxford University Press), 225–258.

Liddell, S. (2003). *Grammar, Gesture, and Meaning in American Sign Language.* Cambridge: Cambridge University Press.

Lindström, J. (2006). Grammar in the service of interaction: exploring turn organization in Swedish. *Res. Lang. Soc. Interact.* 39, 81–117. doi: 10.1207/s15327973rlsi3901_4

Local, J. (1992). "Continuing and restarting," in *The Contextualization of Language*, eds P. Auer and A. di Luzio (Amsterdam: John Benjamins), 273–296.

Local, J., and Kelly, J. (1986). Projection and Ćsilences': notes on phonetic and conversational structure. *Hum. Stud.* 9, 185–204. doi: 10.1007/BF00148126

Martinez, L. (1995). "Turn-taking and eye gaze in sign conversations between deaf Filipinos," in *Sociolinguistics in Deaf Communities*, ed C. Lucas (Washington, DC: Gallaudet University Press), 272–306.

McCleary, L., and Leite, T. A. (2013). Turn-taking in Brazilian sign language: evidence from overlap. *J. Interact. Res. Commun. Disord.* 4, 123–154. doi: 10.1558/jircd.v4i1.123

McIlvenny, P. (1995). "Seeing conversations: analyzing sign language talk," in *Order: Studies on the Social Organisation of Talk and Embodied Activities*, ed P. Ten Have and G. Psathas (Washington, DC: University Press of America), 129–150.

Mehan, H. (1979). *Learning Lessons. Social Organization in the Classroom.* Cambridge: Harvard University Press.

Mondada, L. (2007). Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Dis. Stud.* 9, 194–225. doi: 10.1177/1461445607075346

Oloff, F. (2013). Embodied withdrawal after overlap resolution. *J. Pragmat.* 46, 139–156. doi: 10.1016/j.pragma.2012.07.005

Rasmussen, G., Hazel, S., and Mortensen, K. (2014). *A Body of Resources - CA Studies of Social Conduct: Special Issue for the Journal of Pragmatics*, Vol. 65. Amsterdam: Elsevier B.V.

Rossano, F. (2012). "Gaze in conversation," in *The Handbook of Conversation Analysis*, eds J. Sidnell and T. Stivers (Malden, MA: Wiley-Blackwell), 308–329.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Schegloff, E. A. (1982). "Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences," in *Analyzing Discourse:*

Text and Talk, ed D. Tannen (Washington, DC: Georgetown University Press), 71–93.

Schegloff, E. A. (1996). "Turn organization: one intersection of grammar and interaction," in *Interaction and Grammar*, eds E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 52–133.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29, 1–63. doi: 10.1017/s0047404500001019

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organisation of repair in conversation. *Language* 53, 361–382. doi: 10.1353/lan.1977.0041

Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics* 6, 371–388. doi: 10.1075/prag.6.3.06sel

Selting, M. (2000). The construction of units in conversational talk. *Lang. Soc.* 29, 477–517. doi: 10.1017/S0047404500004012

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Stivers, T., and Rossano, F. (2010). Mobilizing response. *Res. Lang. Soc. Int.* 43, 3–31.

Tannen, D. (1984). *Conversational Style: Analysing Talk Among Friends*. Norwood, NJ: Ablex Publishing Corporation.

Thibeault, A. (1993). "Overlap in Filipino sign language discourse," in *Paper Presented at Communication Forum*, ed E. Winston (Washington, DC: Gallaudet University), 207–218.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 September 2015
doi: 10.3389/fpsyg.2015.01326

Suspending the next turn as a form of repair initiation: evidence from Argentine Sign Language

Elizabeth Manrique[1]* *and N. J. Enfield*[1,2]

[1] Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [2] Department of Linguistics, The University of Sydney, Sydney, NSW, Australia

Practices of other-initiated repair deal with problems of hearing or understanding what another person has said in the fast-moving turn-by-turn flow of conversation. As such, other-initiated repair plays a fundamental role in the maintenance of intersubjectivity in social interaction. This study finds and analyses a special type of other-initiated repair that is used in turn-by-turn conversation in a sign language: Argentine Sign Language (Lengua de Señas Argentina or LSA). We describe a type of response termed a "*freeze-look,*" which occurs when a person has just been asked a direct question: instead of answering the question in the next turn position, the person holds still while looking directly at the questioner. In these cases it is clear that the person is aware of having just been addressed and is not otherwise accounting for their delay in responding (e.g., by displaying a "thinking" face or hesitation, etc.). We find that this behavior functions as a way for an addressee to initiate repair by the person who asked the question. The "freeze-look" results in the questioner "re-doing" their action of asking a question, for example by repeating or rephrasing it. Thus, we argue that the "freeze-look" is a practice for other-initiation of repair. In addition, we argue that it is an "off-record" practice, thus contrasting with known on-record practices such as saying "Huh?" or equivalents. The findings aim to contribute to research on human understanding in everyday turn-by-turn conversation by looking at an understudied sign language, with possible implications for our understanding of visual bodily communication in spoken languages as well.

Keywords: conversation analysis, Argentine Sign Language, visual-gestural modality, "*freeze-look*", other-initiation of repair, questions, responses

OPEN ACCESS

Edited by:
Manuel Carreiras, Basque Center on Cognition, Brain and Language, Spain

Reviewed by:
Marcel Giezen, Basque Center on Cognition, Brain and Language, Spain
Brendan Costello, Basque Center on Cognition, Brain and Language, Spain

***Correspondence:**
Elizabeth Manrique, Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, Netherlands
elizabeth.manrique@mpi.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *13 March 2015*
Accepted: *18 August 2015*
Published: *15 September 2015*

Citation:
Manrique E and Enfield NJ (2015) Suspending the next turn as a form of repair initiation: evidence from Argentine Sign Language. Front. Psychol. 6:1326. doi: 10.3389/fpsyg.2015.01326

Introduction

People in interaction use and interpret meaningful hand and facial gestures spontaneously and frequently as part of their efforts to express themselves and to understand others when formulating turns in conversation. In spoken languages, these visible behaviors form an integrated multimodal system with speech, where the visible and audible signs are linked pragmatically, semantically, and temporally (McNeill, 1992; Kendon, 2004; Enfield, 2009). In sign languages, visible behavior bears the entire load: People rely solely on visual-gestural communication when producing linguistic signs and communicative gestures, coordinating multiple bodily resources including manual, facial, and head signs and movements (Klima and Bellugi, 1979; Emmorey, 2001; Sandler and Lillo-Martin, 2006). But no matter what combination of modalities and semiotic resources is

used, all language users are faced with the challenge of maintaining mutual understanding in the turn-by-turn flow of conversation (Clark, 1996).

Problems of perception or understanding in conversation occur very often, with other-initiation of repair occurring on average around once every 100 s (Dingemanse et al., in press). To understand how these alerts are handled in real time, we must focus on the basic organizational structure of everyday conversation, namely the sophisticated systems of turn-taking (Sacks et al., 1974; Stivers et al., 2009; Levinson and Torreira, 2015) and sequence organization (Clark, 1996; Schegloff, 2007). When a person asks a question, they are taking a turn at talk of the kind that obliges another person to produce an answer or other relevant response in the next turn. Different responses can be displayed. An addressee can simply answer a question directly if that is possible. But if they do not understand or do not hear the question clearly, they have the option of initiating repair by the questioner, for example by saying in English *"Sorry?," "What?," "Huh?,"* or *"Can you repeat that?."* This is called *other-initiation of repair*, abbreviated as OIR (Schegloff et al., 1977; Dingemanse et al., 2013; Dingemanse and Enfield, 2015).

Research on other-initiation of repair to date has been done almost exclusively on spoken languages, in telephone and face-to-face interaction, with a fairly limited sample of languages beyond English (see Dingemanse and Enfield, 2015; Hayashi et al., 2013 for recent crosslinguistic studies). Despite an explosion of recent research on sign language in linguistics and related fields, there is relatively little research on interactional structures and mechanisms in sign languages, especially where such research focuses on naturally-occurring interaction. Available studies deal with aspects of turn-taking in American Sign Language (Baker, 1977), tactile Sweden Sign Language (Mesch, 2001), Brazilian Sign Language (McCleary and Leite, 2013), and Sign Language of the Netherlands (de Vos et al., 2015), as well as repair practices in American Sign Language (Dively, 1998), and Tactile Australian Sign Language (Willoughby et al., 2014). The present study draws on an extensive corpus of videotaped conversation (both dyadic and multi-party) in a sign language, giving extensive access to spontaneously occurring data on repair practices that rely solely on the visual-gestural modality.

Though repair practices have been traditionally defined as dealing with problems of "speaking, hearing, and understanding" (Schegloff et al., 1977), in sign language these must instead be understood as problems of "*signing*, *seeing*, and understanding." Signers use a variety of body articulators in coordinated ways to produce visible linguistic information: these include hand movements, facial expressions, eye gaze, head, and body postures and mouth action signs (Baker, 1977; Baker-Shenk, 1983; Sutton-Spence and Woll, 1999; Boyes-Braem et al., 2001; Liddell, 2003; Sandler and Lillo-Martin, 2006; Vermeerbergen et al., 2007). Other-initiated repair can in principle be produced by any of these articulators or, more commonly, by a combination of them in the repertoire of Argentine Sign Language (LSA) practices (and in other sign languages). Ongoing research on LSA is investigating the full set of types of OIR found in a conversational corpus (Manrique, in press). The "freeze-look" behavior described and analyzed in this article is one of these OIR types.

There are obvious and important differences between the role of the visual modality in spoken vs. signed languages, and it may be expected that these affect the ways in which people encounter and handle problems of perception and understanding. We will distinguish between the *seeing problems* that can lead to other-initiation of repair in sign languages, vs. the *hearing problems* that can occur in spoken languages. Yet there have been recent suggestions of strong commonalities between signed and spoken languages in this domain. Enfield et al. (2013) compare linguistic and conversational mechanisms in relation to problems of understanding in a sample that included LSA and 20 spoken languages across the globe, with results suggesting linguistic and conversational universals in social interaction. LSA signers have the same basic functional options as those described for spoken languages, such as "open" vs. "restricted" formats for other-initiation of repair (Dingemanse and Enfield, 2015; see below for definitions). Another study (Floyd et al., 2014), focusing on the use of "holds" in OIR sequences in LSA and two unrelated spoken languages, Italian and Cha'palaa, also suggests commonalities across signed and spoken languages concerning the function and timing of final-turn holds. In this study visual bodily, including head, face, hands, or torso, or any combination of these components, was compared when initiating repair to another person. This visual bodily behavior is characterized by the maintenance of at least one of these components as strategy of pursuing a resolution of understanding problems. This study has shown that in most of the cases the hold behavior was disengaged only once the person who has initiated repair had heard or seen some or all of the repair solution-turn produced by the person of the trouble source.

If we are going to understand how systems of turn-taking are managed in real time, it is crucial to understand how problems of perceiving or understanding are dealt with on the spot. After all, given the fast pace and constant forward progression of turn-by-turn conversation, if a problem is not fixed immediately then the chance to fix it may quickly be lost. This paper provides a perspective from sign language analysis with the broader aim to gain a better understanding of the general phenomenon of repair as a back-up mechanism for possible threats to the collaborative progress of conversation.

Other initiated-repair occurs necessarily in dialogue, and specifically within the context of conversational turn-taking. A basic OIR sequence has three turn elements (Dingemanse et al., in press). The center or pivotal point in the sequence is the *initiation of repair* (here referred to as T0). T0 points back to the previous turn and identifies it as problematic in some way. This previous turn is termed the *trouble source* of the sequence (referred to as T-1). Usually, T0 explicitly asks that T-1 should be fixed in some way: examples are *Huh?*, *What?*, *Who?*. Following T0 is the *repair solution* (or T+1), produced by the person who produced the original trouble source turn. We refer to the producer of the trouble source and repair solution as Person A, and the person who initiates the repair sequence as Person B. If the repair solution by Person A is not sufficient to solve the problem, then Person B might pursue with another initiation of repair asking for more clarification or repetition, thus expanding the sequence. Once B is satisfied with the solution he or she may provide an

uptake turn (T+2) indicating or at least claiming that he/she has now satisfactorily heard/seen or understood what was said (Schegloff et al., 1977; Clark, 1996).

Our focal point of interest here is the linguistic format of T0, the nuclear turn of the sequence, in which other-initiation of repair is done. T0 turns can display different forms to indicate different problems of perception or understanding in spoken language. Two macro categories of OIR are defined by how they specify the scope of the problem that Person B is targeting in the previous turn. These are "open" and "restricted" categories of OIR (Dingemanse and Enfield, 2015). Open type repair initiators do not specify what the problem is or where it is located in the previous turn produced by Person A: examples of open type repair initiators include *Huh?* and *Pardon?*. These repair initiators point to the entire previous turn as problematic. By contrast, restricted type repair initiators specify what the problem is and where it is located: examples include *Who?* and *They said what?*. These specific types of repair initiators limit the scope of the problem, indicating that the problem is not with the entire previous turn but a part of it.

Previous research on other-initiation of repair has focused on *explicit* or *on-record* ways of initiating repair on other participants' turns. If a speech act is on-record this means that it is non-deniable. For example, if one makes a threat in on-record form (e.g., "If you don't pay up I will hurt your family"), then one would be unable to plausibly deny (say, in court) that it had been a threat. By contrast, a communicative act is done *off*-record "if it is done in such a way that it is not possible to attribute only one clear communicative intention to the act" (Brown and Levinson, 1987, p. 211). An off-record strategy (such as "You should probably pay up. By the way, how is your family? It would terrible if something happened to them") might be obvious in its communicative intention and yet that intention would be plausibly if not at least technically deniable. Off-record strategies are typically used when people want to avoid possible consequences of being held to account for having performed certain social actions. This is sometimes for legal reasons as in the case of the threat, or perhaps more often it is a way of minimizing the "face-threatening" nature of many types of speech act (Brown and Levinson, 1987). If someone does a communicative act in an off-record way, they are technically leaving it open to the other person to decide how to interpret that act.

In the domain of other-initiation of repair, if a person says "Huh?" or similar known OIR strategy then they are initiating repair in an on-record way. They would be unable to deny that they had intended to momentarily suspend the progress of the conversation in order to resolve a problem of perception or understanding. Here we aim to expand current knowledge of OIR systems by describing a systematic *off*-record practice for initiating repair. This is the "freeze-look," observed here in LSA: a question is posed, but this question is a source of trouble for the one who is required to provide an answer; rather than providing an answer, the addressee produces a *freeze-look*, meaning that they hold their body and manual articulators still while gazing directly at their interlocutor. In these cases signers continue looking at the questioner without giving any signal that an answer is coming soon. They do not move, and are thus not visibly "gearing up" to respond. We find that the questioner typically treats this practice in the same way as they would treat an open format of other-initiated repair (such as "Huh?"), namely, by repeating or rephrasing the question. We argue that this "freeze-look" behavior is a dedicated but off-record practice for open other-initiation of repair in LSA. The practice allows us to distinguish between on-record OIR and off-record OIR practices in visual-gestural modality in a sign language, and it suggests a distinction in OIR strategies that might be found in other languages, including spoken languages.

Argentine Sign Language (LSA)

LSA is used in Argentina, mostly in the city of Buenos Aires, Greater Buenos Aires, Cordoba, and Mendoza. According to the last official report (INDEC[1]) in 2010, there are 289,321 hearing-impaired people in Argentina out of a total population of the country of 41.499 million people. However, there are no official surveys regarding LSA users. LSA is influenced in some ways by contact with Spanish, for example in the common use of Spanish words, either mouthed or fingerspelled. Members of the LSA community vary with respect to their background: a small minority are deaf with deaf parents, most are deaf with hearing parents, others are hearing but have learned to use the language, for example because their parents or other family members are deaf. Beyond the schooling system, deaf clubs and associations provide a context in which LSA is used and learned.

LSA is historically related to Italian Sign Language (Veinberg, 1996). Previous work on the language includes mainly description of the grammar (Massone and Machado, 1992; Massone and Curiel, 1993; Curiel and Massone, 2004), dictionaries (Massone, 1993; Valassina, 1997), and work on deaf bilingual education, interpretation, and other issues (Behares et al., 1990; Veinberg, 1996).

Data and Method

The LSA data used in this study were sampled from recordings of everyday informal dyadic and multi-party interactions made in Deaf clubs and Associations in Buenos Aires, Argentina. LSA users usually meet in these places to interact and share social, educational, sport and political activities. The recordings were filmed without modifying the natural and daily environment of the signers where they normally carry out their activities. The recordings were done as unobtrusively as possible. Participants were not given any instructions or tasks to perform. All participants were native LSA[2] adult friends. Both men and women took part. The materials were collected with fully informed consent under formal ethics clearance approved by the funding body (European Research Council) and the

[1] Available online at: http://www.censo2010.indec.gov.ar/resultadosdefinitivos_totalpais.asp.

[2] We classify native signers as proficient individuals that have been exposed to a sign language, in this case LSA, from birth or infancy from their caregivers.

host institution (Max Planck Society), and also in line with ethical guidelines of the DOBES program (Documentation of Endangered Languages). All the videos were recorded in institutional settings where the relevant authorities authorized the recordings in advance. All participants were informed about the purposes of the research (namely, the study of language use in naturally occurring interactions) and all gave consent before being filmed. Participants signed informed consent statements that provided information about the study, the researchers, and the institutions responsible. They gave permission for the data to be used for research and educational purposes including academic and educational publications. The informed consent forms were written in Spanish and were also translated into LSA by the first author (a certified interpreter in LSA), who collected the data.

The video corpus was collected by the first author between 2010 and 2012. It was filmed using two high definition cameras (Canon HDV). For this study, a sample was taken from the larger corpus by selecting segments of between 10 and 20 min from different recordings to ensure a variety of interactions and participants, totalling 1 h and 50 min of conversation. In this selected part of the corpus 59 signers have participated, between 20 and 80 years old, 35 men and 24 women. Two hundred and thirteen cases were collected to form a set of cases of other-initiation of repair (OIR) for a large-scale comparative research project (see Dingemanse and Enfield, 2015; Manrique, in press). The cases collected were transcribed and translated in collaboration with native signer consultants. From this set, 10% (23 out of the 213) were identified as cases of the "freeze-look" behavior we focus on here. 23 signers, 15 men and 8 women, between 20 and 65 years old have participated in this smaller collection of examples. These "freeze-look" cases were transcribed, glossed, annotated, and translated into English in the transcription software ELAN (Wittenburg et al., 2006). The transcription consisted of sign-by-sign translation into Spanish[3] following the original sign order, done in collaboration with native LSA consultants. These were annotated using sign language glosses and linguistic notational conventions based on the Johnston (2010) system for sign languages, drawing also on Jefferson's spoken language conversation notation system (Jefferson, 2004, 2015), with some innovations specific to LSA (Manrique, 2011). The examples collected for the OIR collection were translated into English.

Coding

The basis for identifying and coding the "freeze-look" behavior for this study include formal criteria of the behavior, and distributional criteria in terms of the conversational sequence in which the behavior occurs. The formal criterion of the "freeze-look" action itself is that the body is held still and the gaze is directed straight at the other person. This alone is not enough, though: there is also a distributional criterion for this study, namely that the behavior occurs immediately after a question by the other person in a conversation. The "freeze-look" cases were identified for this study in the context of a larger study of other-initiation of repair (OIR) in LSA and other languages. Subsequent sections provide the details on how the cases were identified and coded.

Identification and Coding of Other-initiated Repair (OIR)

A coding system for OIR was created as part of a major comparative project of video-recorded corpora in 12 languages, including LSA (see Dingemanse et al., in press for a detailed description of the coding schema). The design of the coding system was based on observations of conversational data, taking into account extensive prior work on OIR, mainly in spoken English, and enriched with special attention to cross-linguistic diversity and multimodal information.

Sequences of other-initiation of repair in LSA were identified and annotated, using multiple tiers in ELAN to code information about grammatical, pragmatic and sequential properties of each case. Independent tiers were created per participant to annotate grammatical and pragmatic information including independent tiers for signers' right and left hands, and for non-manual markers including: eyebrows (raised, together), eyes (wide open, squint, closed), eye gaze, wrinkled nose, mouth gestures, mouthing, head movements, and upper-body movements. Each example was identified with a unique ID, and the three core turns of each sequence were distinguished: (1) trouble source (e.g., A: *Have you seen John?* = T-1), (2) initiation of repair (e.g., B: *Who?* = T0), and (3) solution turn (e.g., A: *John.* = T+1), along with an "uptake" or sequence-closing turn if relevant (e.g., B: *Ah, no, I haven't seen him today.* = T+2). **Table 1** shows the distribution and frequency of types of OIR cases in the collection (Manrique, in press). In LSA, restricted type repair initiators are nearly twice as frequent as the open type.

Coding and Transcription of "Freeze-look" Cases

"Freeze-look" cases were coded for numerous features, including timing aspects and formal aspects. Three measures of timing of "freeze-look" cases were coded on independent tiers in the annotation software. These timing measures were as follows:

(a) the length of the entire sequence (T-1, T0, and T+1);
(b) the duration of the "freeze-look" (see Section Timing of "Freeze-looks" below);
(c) the time between the end of the question (T-1) and the beginning of the (near) repeat of the question (T+1), both produced by *A* (see Section Timing of "Freeze-looks" below).

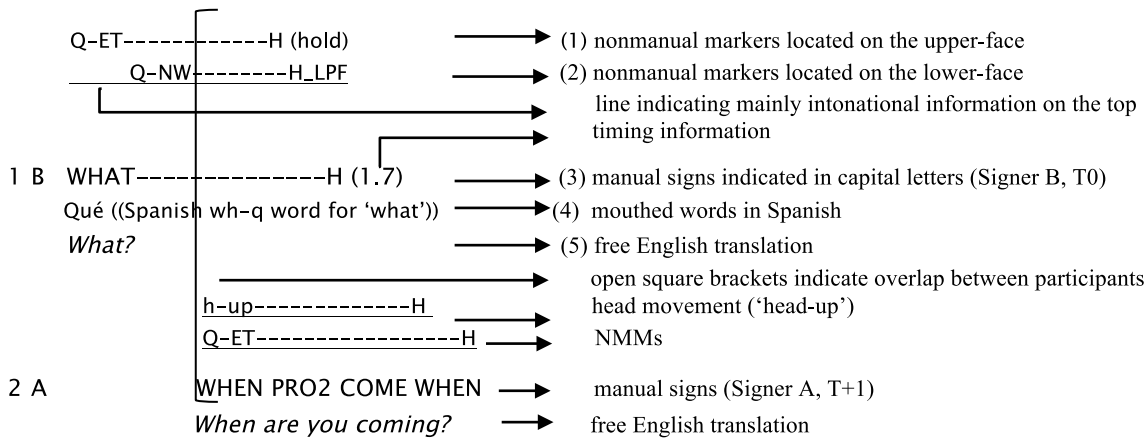
Formal coding of the "freeze-look" cases included the information that we provide in the data examples below, presented using between a minimum of one line and a maximum of five lines. Two lines are distinguished for non-manual markers (NMM), the first one for head movements (e.g., head-down) and the second one for facial movements (e.g., ET "eyebrows together"). The line below NMM information (see schema below) indicates the extension and alignment of NMM (above

[3]The data were transcribed in written Spanish because LSA does not have a written system for this purpose. Most of the signers are bilingual in spoken Spanish and LSA. They use Argentinian Spanish in everyday life for fingerspelling, mouthing, speaking, writing and reading.

TABLE 1 | Frequency of the types of repair initiators in the LSA corpus (Manrique, in press).

Type		Subtype	Frequency (n/213)	Proportion (%)
Explicit (on-record)	Open	Non-manual[4]	51	23
		Question-word (*What?*)	13	6
		Formulaic	0	0
	Restricted	Content q-word (asking for specification)	15	6
		Repetition (asking for confirmation)	56	26
		Offer (asking for clarification)	52	24
		Alternative question	3	1
Implicit (off-record)	Open	"Freeze-look" response	23	10

the line) in relation to manual makers (below the line) that are often produced in overlap. In general, one line is used for manual sign glosses (mainly lexical information, illustrated in line 3 below) giving single-word translations into English in capital letters. In some examples it is also relevant to include more information to indicate distinctive use of the separate hands. In these cases, one line is used for the right hand and another one for the left hand. Mouthing is also used frequently in OIR practices and it is indicated by a separate line (see line 4) after the manual glosses when it is relevant. The last line corresponds to the free English translation in italics. Here is an example, illustrating the distinct lines for representation of each of the formal aspects that we coded.



A *large open bracket* indicates when overlapping turns are produced between participants. At the end of line (3), the timing information of the duration of a sign is indicated between parentheses in seconds (1.7). In line (4), double parentheses contain additional comments from the transcriber (see Supplementary Material for a full description of conventions used in the examples in this article). Besides the transcription of the cases described above, a summary of every example is also provided for easier access to the data, including only the main OIR sequence (T-1, T0, and T+1) in free English translation.

Results

Question-answer Sequences

This study focuses on question-answer sequences in unscripted sign language interaction. Question-answer sequences are one type of *adjacency pair* (Schegloff, 1968; Schegloff and Sacks, 1973; Schegloff, 2007). In an adjacency pair, an initial move by Person A creates a normative requirement for Person B to produce a response, where that response is expected to be of a particular kind. For example, in a question-answer adjacency pair, the first *pair part* (e.g., "What time is it?") puts the other person in a position where they are obliged to respond appropriately. The preferred response to a question is an *answer* (e.g., "9 o'clock"), but the normative obligation to deal with the question can also be handled by a *non-answer response* that is still relevant to the question (e.g., "Sorry, I don't have a watch"). Both answers and non-answer responses are adequate as second-pair parts to questions. There are, however, more vague or ambiguous types of things one might do immediately after a question. For example, one could stay silent and not move. This could of course be taken as a complete lack of response, if for example Person B did not

[4]They include eyebrow actions (raised and together), wrinkled nose, lips puckered forward, open-mouth gestures (resembling "*Huh?*" in spoken language), head movements (up, down, side, backwards), and leaning forward (Manrique, in press).

realize that Person A was talking to them at all. But it could also be taken as a specific way for Person B to provide a non-answer response, not just a failure to respond but a way of signaling that one is not going to respond. This is the possibility we explore in subsequent sections.

Before proceeding, we briefly describe how questions are formed in LSA. Signers in LSA mark questions with non-manual markers (Veinberg, 1993) as also is often the case in other sign languages (Baker and Padden, 1978; Baker-Shenk, 1983; Sandler et al., 2011). The use and the timing of non-manuals are coordinated and linguistically constrained to the manual sign(s) with which they co-occur (e.g., Baker-Shenk, 1983; Pfau and Quer, 2010). The main non-manual markers for questions in LSA are *eyebrows together* for WH-questions ("What?," "Who?" etc.) and *eyebrows raised* for yes/no questions. These eyebrow positions can be combined with head upward or downward movements and/or upper body leaning forward. Another important characteristic of questions in LSA is the presence of eye gaze directed from the person who asks the question to the addressee of the question. It occurs during and after the question has been produced. At the end of the question, the questioner usually then momentarily suspends or "holds" at least one of the elements that compose the question. These can be manual signs (e.g., *"What?," "Who?,"* etc.), or non-manual components such as facial actions that indicate that a question has been produced.

Fitted Responses to Questions

When Person A asks Person B a question, Person A ideally expects an answer: this would "fit" best as a response. This is illustrated in the next example. In Extract 1 Signer A asks a question about Signer B's children (line 1) and Signer B immediately provides a fitted answer in the following turn (line 2).

Summary: Extract 1

1 A Are your two children OK?
2 B Good, very good, very good

Extract 1. SM | 161990

		Q-ER	
1	A	CHILDREN TWO GOOD=	((Question))
		Are your two children OK?	
2	B	=GOOD VERY-GOOD VERY-GOOD	((Fitted answer))
		Good, very good, very good.	

Non-fitted Response

There are numerous ways in which someone might produce a *non-fitted response* to a question; i.e., something other than an answer. In this section we discuss three types of non-fitted response—non-attendance, word search, and on-record repair—before giving closer attention to a fourth type of non-fitted response, which is the focus of this study.

Non-response due to non-attendance

One way of producing something other than a fitted response is not to respond at all. In sign language this can happen if the addressee was not looking at the other person when the question was asked, or if they were interrupted or distracted by someone or something else when the question was asked. In these cases, there is an obvious account for why no response is given: it is clear in the situation that the question was not properly attended to and could not have been perceived or understood. In this situation, the questioner needs to secure the addressee's attention before redoing the question.

The next example shows numerous strategies to get another signer's attention when they are clearly not perceiving what is being signed to them (cf. Baker, 1977). These include directing the eye gaze, tapping the addressee and holding the hands up waiting for the addressee's attention. In the example given below, Signer A starts asking B a question (line 1), but at that moment Signer B is signing and looking at another person, C. Signer A tries again by holding Signer B's arm to get her attention, but B continues signing to C. Then, Signer A maintains the last manual sign she has produced still while looking at B and waiting for her attention. In lines 3 and 4, A re-initiates the question when B (line 4) turns her head toward A. However, A has her eyes closed and looks back to C again. In line 5, A tries again to get B's attention by holding B's arm, B looks at A and A repeats the question (*"Sunset?"*) adding the sign *"time?"* to finish the question *"What time is the sunset (there)?"* This refers to the sunset in a different region in Argentina (Perito Moreno Glacier in the South). In line 8, B answers A's question after several attempts from A. In this example it is clear that the lack of response from B to A's initial attempts to ask her question is due to non-attendance and failure to perceive what was being signed.

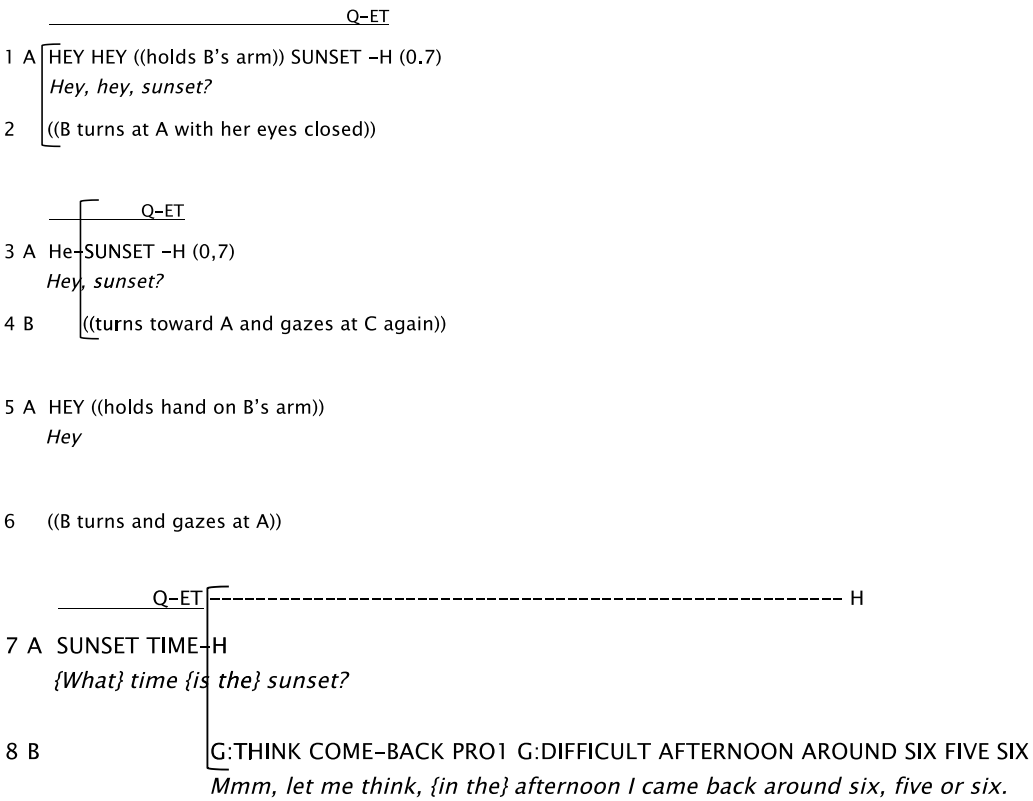
Extract 3 shows a similar example in a dyadic interaction between two friends. Although one of the participants (B) is signing to A in this example, he is not maintaining eye contact all the time, but shifting eye gaze, closing his eyes, looking at his hands while occasionally monitoring his addressee. In line 1, when Signer A asks Signer B a question, Signer B is not looking at A. Then, Signer A maintains both hands in signing position, pointing at Signer B (*"PRO2"*), waiting for his attention. In line 3, once B opens his eyes and looks at A, then A repeats the question. B recognizes A is asking a question, drops his last sign and answers A (*"NO, I haven't sent it to the office"*), followed by a clarification, *"I sent it, but I haven't read it."*

Visual contact and feedback play different roles in signed and spoken language conversation (see Baker, 1977). Constant visual feedback and mutual monitoring between parties in sign language conversation is indispensible to successful communication. Signers generally maintain more focused interactions and minimize multi-tasking activities that would divert visual attention from the interaction. Signers provide constant feedback and monitoring using manual and nonmanual attention-getting strategies (Baker, 1977). More research is needed to determine how different signed and spoken language everyday conversation are in this regard.

Summary: Extract 2

1	A	Hey, hey, sunset?	((**Initiating a question**))
2	B	((Not looking at A))	((no visual attention))
3	A	Hey, sunset	((re-initiating the beginning of a question produced at line 1))
4	B	((Not looking at A))	((no visual attention))
5	A	Hey, ((holds hand on B's arm))	((uses attention getting strategy))
6		((B turns and gazes at A))	((visual attention))
7	A	{What} time {is the} sunset?	((re-initiating the question that A started at line 1 and 3))
8	B	Mmm, let me think, {in the} afternoon I came back around six, five or six.	((**Answer**))

Extract 2. MS | 1708720



Non-fitted response with signs of "word search"

A second way of giving a non-fitted response to a question is to give an explicit signal that the response is delayed due to inability to find the words one is looking for in formulating an answer. Word-searching displays are common type of non-fitted response in everyday interaction. They indicate that the addressee is working on the answer and that the answer is delayed. Speakers use different vocal and gestural strategies to indicate they are working on the answer such as: cutoffs, fillers ("*um,*" "*uh,*" etc.) (Levelt, 1983; Clark and Fox Tree, 2002), and break of eye contact (often then looking upwards) (Goodwin and Goodwin, 1986). Signers use similar gestural visual strategies to indicate problems in delivering or remembering a specific reference, including shifting eye gaze, closing eyes, "thinking" gestures, or rubbing fingers.

Extract 4 shows multiple word searching strategies by both participants after a question is asked. These are: breaking eye contact, squinted eyes (line 4); giving an *ad-hoc* description instead of name ("short hair"), holding hand/s up, and closing eyes (line 5) when trying to retrieve a name using fingerspelling (line 8).

In examples like this one, a signer's observable "thinking" behavior is a way of overtly accounting for the failure to provide an answer to a question. As it also indicates that the signer has indeed understood the question, it does not elicit a repetition of the question.

Non-fitted response: on-record repair

A type of non-fitted response that is always possible is an explicit, on-record other-initiation of repair (Schegloff et al., 1977;

Summary: Extract 3

1	A	Wait, what's your topic?	((**Initiating a question**))
2	B	**((Signing, not looking at A))**	((no visual attention))
3	A	What's your topic?	((repeat of the question))
4	B	**((Looks at A))**	((visual attention))
5	A	No, I haven't sent it to the office.	((**Answer**))
6	B	Ah, OK	((information up-take))

Extract 3. JP | 234070

```
                                h-up
                               Q-ET
1 A  WAIT PRO2 TOPIC PRO2-H
     Wait, what's your topic?
2 B  ((Signing, not looking at A))

     Q-   ET h-up
3 A  TOPIC PRO2------H
     What's your topic?
4 B  ((Looks at A))
5 B  NO PRO1 SEND-STUDY NO PRO1
     No, I haven't sent it to the office.
(...)
     ER       nod
6 A  PT-PRO2-H
     Ah, OK.
```

Schegloff, 1982; Hayashi et al., 2013; Dingemanse and Enfield, 2015). If one has not heard or understood a question, it is always possible to ask for repetition or clarification of the question rather than attempt to answer it. Other-initiation of repair is a way of dealing with online problems of hearing and understanding during interaction so as to maintain and secure mutual understanding, alignment, and affiliation. It is, however, dispreferred, as it halts the progress of talk during a conversation, derailing it momentarily (Stivers and Robinson, 2006).

Extract 5 shows an explicit initiation of a repair sequence after a question as illustrated in **Figure 1** by Signer B. It is done using an "open" format ("What?") in line 2. There is also a "restricted" type of repair initiation ("Inside?") by Signer A, in line 5.

In this example, Signer B displays an explicit initiation of repair using both manual signs (WH-question word "*WHAT*") and nonmanuals (eyebrows together and leaning forward). Signer B holds these until Signer A solves the problem by near-repeating the trouble source (i.e., the question). In the solution turn, Signer B makes the implicit question more explicit, and more specific (by clarifying the person referent, "*Your son, the-oldest*"). Another initiation of repair in this example is done using a restricted format (line 5; for the terminology "restricted" vs. "open," see Section Introduction, above, and Dingemanse and Enfield, 2015). It is produced with a combination of manual markers (the hand sign for "INSIDE") and nonmanual markers of yes/no questioning (raised eyebrows and head moving downwards).

Freeze-look: A Notable Absence of Response

We now turn to the type of non-fitted response that we refer to as a *freeze-look*. We argue that this type of response is a non-official or off-record way of initiating repair. In a collection of cases of other-initiated repair in LSA (Manrique, in press), the "freeze-look" practice makes up around 10% of all cases. The "freeze-look," which effectively prompts a questioner to re-do their question, is performed by an addressee by holding their hands and body in position and looking directly at the questioner at a time when it is expected that they should be now responding to the question. This suspended or frozen body posture is maintained until the signer of the trouble source redoes the question (e.g., by repeating or rephrasing), or until the person producing the "freeze-look" upgrades by initiating an on-record other-initiation of repair (see below).

The definitive characteristics of the "freeze-look" are the following:

(a) At the relevant moment, the addressee of a question (Signer B) is normatively required to produce a relevant response (an answer to a question, or something related).
(b) The addressee looks directly at the Signer of the question (Signer A).
(c) The addressee temporarily holds their entire body posture in a still or "frozen" position.

Summary: Extract 4

1	A	Hey, {do you} remember a person in primary school ASO,[5] famous?	**((initiating a question))**
2	B	Aha, mm	((hesitation))
3	A	Do you know him/her?	((pursuing the question))
4	B	{Ehhh?} What's his/her name?	((asking for more information))
5	A	Short hair, what was the name? P-e- I can't remember.	((providing extra information))
6	B	Ah, yes, yes, now he's the lawyer here I think	((answer-first attempt))
7	A	No, no, he is from ASO.	((rejection of B's answer))
8	B	**((Closing eyes, moving hands-'thinking gesture'))**	((word searching strategy))
9	A	**((holding hands up and looking at B))**	((pursuing the question))
10	B	C- Pedro ((using fingerspelling and mouthing))	**((Answer))**
11	A	Exactly	((confirmation))
12	B	Yes, yes.	

Extract 4. TFD | 1331595

1 A HEY KNOW TIME-AGO PRIMARY-SCHOOL A-S-O FAMOUS-
Hey, {do you} remember a person in primary school ASO, famous?

2 B -nod++
Aha, mm

Q-ET

3 A PRO2 KNOW PRO2----------H (0.8)
Do you know him/her?

h-up

4 B ((Looks-up, squinted eyes NSP[6] (5.9))) Name?
What is the name? ((mouthing))
{Ehhh?} What's his/her name?

((closing-eyes))

5 A SHORT-HAIR NAME-H P-E- FORGET
Short hair, what was the name? P-e- I can't remember.

h-backwards

6 B YES++ HERE++ LAWYER SEEM LAWYER-H
Mth:a-H now++ lawyer
Ah, yes, yes, now he's the lawyer here I think.

7 A NO + ASO
No, no, he is from ASO.

8 B **((Closing eyes, moving hands-'thinking gesture'))**

9 A ((holding hands up and looking at B))

((Closed-eyes---------H Opened-eyes))

10 B C- P-E- P-E-D-R-O/pedro
C- Pedro ((using fingerspelling and mouthing))

11 A EXACTLY ((Touching B's arm))
Exactly

12 B nod++
Yes, yes.

[5]ASO is a deaf association in Buenos Aires.

[6]NSP: non-signing position.

Summary: Extract 5

1	A	**I** haven't seen him anymore, he disappeared, he travels a lot	T-1
2	B	**What?** **((Figure 1))**	**T0**
3	A	Your son, the oldest, he has left?	T1
4	B	{The} oldest works {at a} ship, {he} works {at a} ship	
5	A	**I**nside?	
6	B	**I**nside, yes, yes	

Extract 5. MS | 161970

		Q-ET	
1	A	WHY **OLDEST** LEAVE-PAST SEE DISAPPEAR LEAVE LEAVE LEAVE=	((Question))
		Why did the oldest leave?	
		I haven't seen him anymore, he disappeared, he travels a lot	
		Q-ET	
2	B	**=WHAT----------H** **((Figure 1))**	**((on-record OIR))**
		What?	
		Q-ET nod	
3	A	-PRO2 SON OLDEST [LEAVE:PAST G:surprise	((near-repeat))
		Your son, the oldest, he has left?	
4	B	[OLDEST SHIP SHIP WORK	((confirmation))
		{The} oldest works {at a} ship, {he} works {at a} ship	
		Q-ER	
5	A	[INSIDE INSIDE	
		Inside?	
		nod	
6	B	[INSIDE INSIDE Mouth:YES YES	
		Inside, yes, yes	

(d) It is clear that the addressee has seen that they were just addressed by A; and they are not otherwise signaling any difficulty in responding.
(e) Signer A then redoes the question (e.g., by repeating or rephrasing).

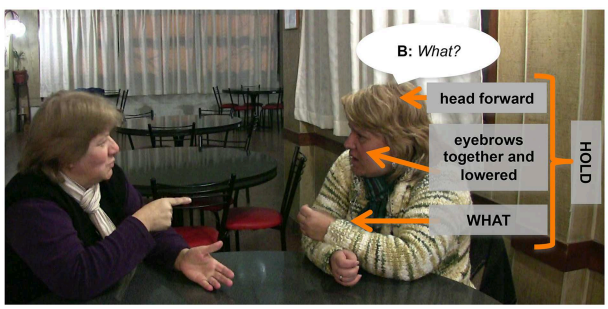


FIGURE 1 | *"What?,"* Signer B, on the right, initiates an open type of repair on A's prior turn in line 2, producing a manual sign for WH-q word *("WHAT")* and nonmanual components (bringing her eyebrows together and leaning forward).

(See diagram in **Figure 2** for a representation of the "freeze-look" response sequence).

We argue that the "freeze-look" is an implicit or off-record practice for initiating repair. Other ways of initiating repair such as asking "What?" are on-record because they use symbolic means to explicitly state that there is a problem of perception or understanding and that this problem now needs to be fixed; the speaker is "officially" committing to their intention to

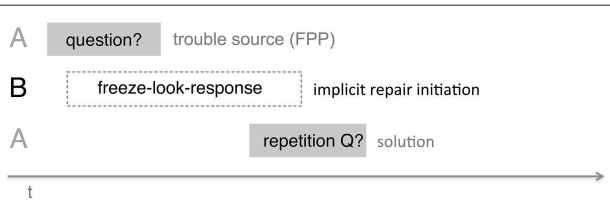


FIGURE 2 | The typical "freeze-look" response sequence, including the question produced by A, as trouble source and first pair part (FPP), the "freeze-look" as a noticeable absence of response turn and the (near) repeat of the question as solution turn provided by the person who earlier produced the trouble source.

momentarily suspend the progress of the interaction, in spite of possible negative or "face-threatening" effects of this (Brown and Levinson, 1987). By contrast, the "freeze-look" is off-record because it does not explicitly encode the intention to initiate repair, just as "It's cold in here" does not explicitly encode the intention to get somebody to shut the window. Nevertheless, as we argue below, the "freeze-look" is highly effective as an other-initiator of repair, but it still technically leaves the interpretation open, thus making a display of giving the recipient some freedom to decide how to interpret the utterance (Sifianou, 1999).

We now present examples of the "freeze-look" phenomenon. In Extract 6, Extract 7, Extract 8, and Extract 9, Signer B produces a "freeze-look" after Signer A has asked them a question. Signer B suspends her/his signing body posture, maintaining it still from the beginning of Signer A's question until near the end of the re-doing of the question as depicted in **Figure 3**. The key point we wish to make here is that in all these cases Signer A treats B's "freeze-look" behavior in the same way as they would treat an explicit open format of other-initiation of repair, namely by immediately re-doing the question (with or without some adjustment). In all these cases, once the question is re-done, B can then produce a fitted response.

In Extract 6, Signer A responds to Signer B's "freeze-look" with a slightly modified repetition of their question, changing the order of the utterance followed by holding his palms up at the end "*eh?,*" and adding more information by specifying a place name ("*Santa-Fe,*" a province in Argentina) (see **Figure 4**).

Summary: Extract 6

1	A	Hey, I was wondering, has (sign-name) moved?	T-1
2	B	**((‘Freeze-look’-response))**	**((Figure 3)) T0**
3	A	I was wondering, has (sign-name) moved to Santa Fe?	**((Figure 4))** T+1

Extract 6. CH-LR | 57000

			ET-G:MM	
1	A	HEY	PRO1-ASK SN MOVE	
		T-1		
		Hey, I was wondering, has (sign-name) moved?		
				Duration in seconds
2	B		**((‘Freeze-look’ response)) (1.9)**	**((Figure 3)) T0**
			ET-G:MM	
3	A	PRO1-ASK SANTA-FE MOVE-		**((Figure 4))** T+1
		I was wondering, has (sign-name) moved to Santa Fe?		

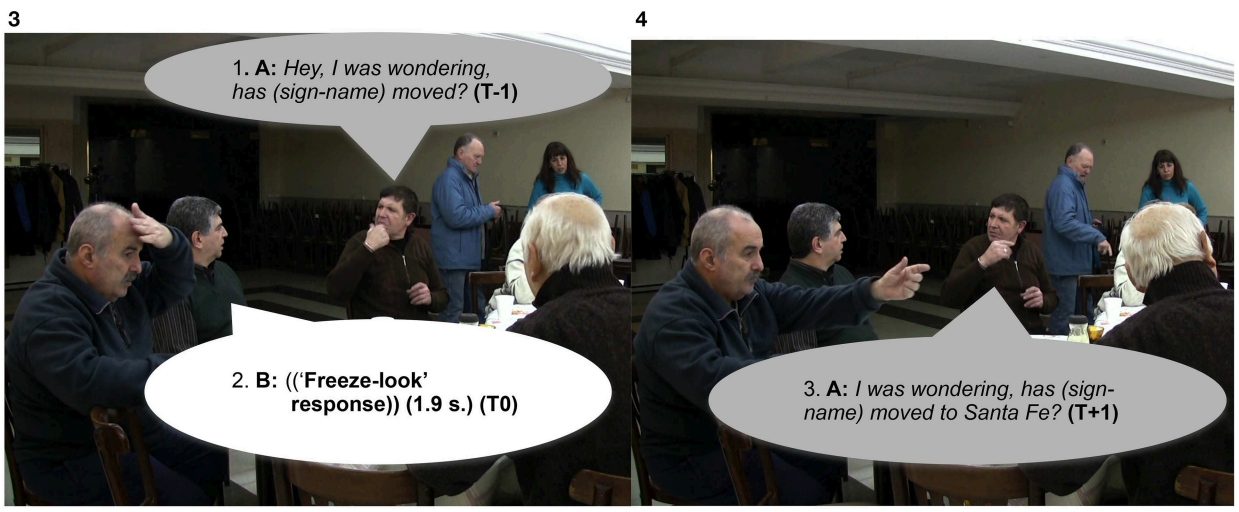


FIGURE 3 | *"Hey, I was wondering, has (sign-name) moved?,"* Signer A, sitting on the right, asks a question to Signer B, sitting on the left (line1). Signer B produces a "freeze-look" response instead of a fitted answer (line 2) that last for 1.6 s.

FIGURE 4 | *"I was wondering, has {sign-name} moved to Santa Fe?,"* Signer A, repeats the question (line 3) produced in line 1 by modifying the order of the utterance and adding more information by specifying a place name.

In Extract 7, Signer B produces a "freeze-look" response to A's question, holding still both manual and nonmanual signs for 1.6 s from the beginning of the question until the end of the re-doing of the question represented in **Figures 5**, **6**. Signer A responds to Signer B's "freeze-look" in the same way they would respond to an on-record OIR strategy, namely by immediately offering a repeat of the question.

In Extract 8, two friends are chatting about vacations in the Perito Moreno Glacier in the South of Argentina. Signer A has vacationed there and Signer B is planning to visit. There is a seeing problem produced by an overlap: both participants are signing at the same time. B's "freeze-look" occurs when Signer A asks the question again: Signer B stops signing, maintaining her nonmanual configuration illustrated in **Figure 7**. The eventual response from Signer A is a re-doing of the question in line 6: a partial repetition, with a change in the order of signs in the utterance shown in **Figure 8**. Signer B then produces a fitted answer (in line 7) as the repeated question is coming to an end.

Summary: Extract 7

1	A	{Has} he taken {something from} you?	**((Figure 5))**	T−1
2	B	**((‘Freeze-look’-response))**		**T0**
3	A	{Has} he taken {something from} you?	**((Figure 6))**	T+1

Extract 7. TFE | 278908

		h-down Q-ER		
1	A	TAKE PRO3 TAKE PRO2 *{Has} he taken {something from} you?*	**((Figure 5))**	T−1
2	B	**((‘Freeze-look’ response))**	**(1.6)**	**T0**
		h-down Q-ER		
3	A	TAKE PRO3 TAKE PRO2 *{Has} he taken {something from} you?*	**((Figure 6))**	T+1

((continues in a 'freeze-look' body position until the end of the repeat of the question line 3))

4	B	GIVE PAY-MONEY PU---H *He has paid, {I don't know}.*	
		nod	
5	A	eyes-blink *Ah, Ok.*	

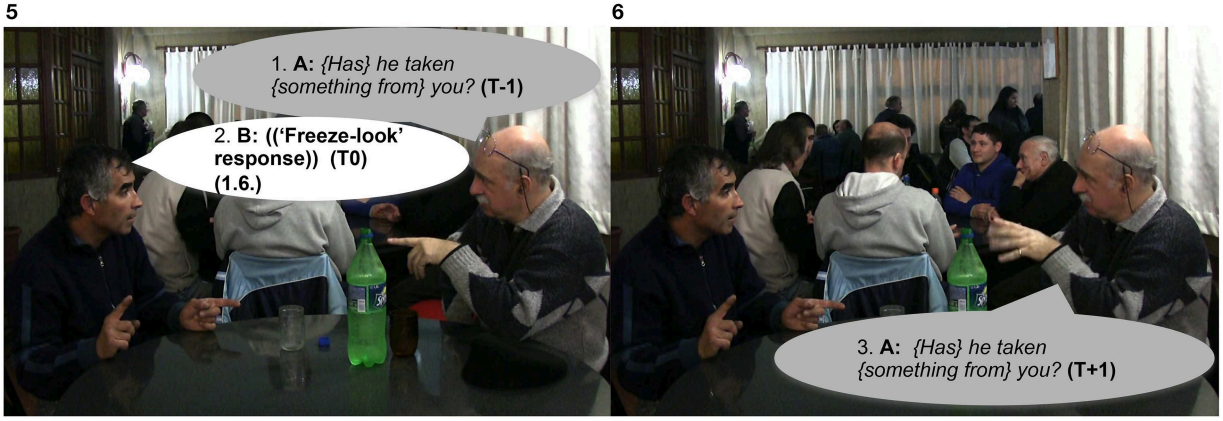


FIGURE 5 | *"{Has} he taken {something from} you?,"* Signer A, sitting on the right, asks B a question (line 1). Then, Signer B, sitting on the left, displays a "freeze-look" response for 1.6 s (line 2).

FIGURE 6 | *"{Has} he taken {something from} you?,"* Signer A, sitting on the right, repeats the question to B (line 3). In parallel, Signer B, sitting on the left, continues with the same "freeze-look" until Signer A finishes the repetition of the question (line 3) and then answers the question.

Summary: Extract 8

4	A	Haven't **I** showed you the pictures?	**((Figure 7))**	T−1
5	B	**((‘Freeze-look’ response))**		**T0**
6	A	The pictures, have **I** showed you them?	**((Figure 8))**	T+1

Extract 8. MS | 919570

1	A	VERY-COLD VERY-COLD SEE-PRO3(glaciers) VERY-CLOSE	
		It's very cold, but you can see (the glaciers) very close	
2		PICTURE PRO1-SHOW-PRO2----------H NOTHING?] ((Seeing problem))	
		Haven't I showed you the pictures?	
3	B	CLOSE----------H?	
		(Can you see them) close?	
4	A	PRO1-SHOW-PRO2 PICTURE-H (0.x) picture?((Holding ‘PICTURE’))	T−1
		Haven't I showed you the pictures? **((Figure 7))**	
5	B	**((‘Freeze-look’ response))** **(1.4)**	**T0**
6	A	PICTURE PRO1-SHOW-PRO2-PRO1-H? **((Figure 8))**	T+1
		The pictures, have I showed you them?	
7	B	YES YES	T+2
		Yes, yes	
8	A	SEE-PRO3p VERY-CLOSE	
		You can see them very close.	

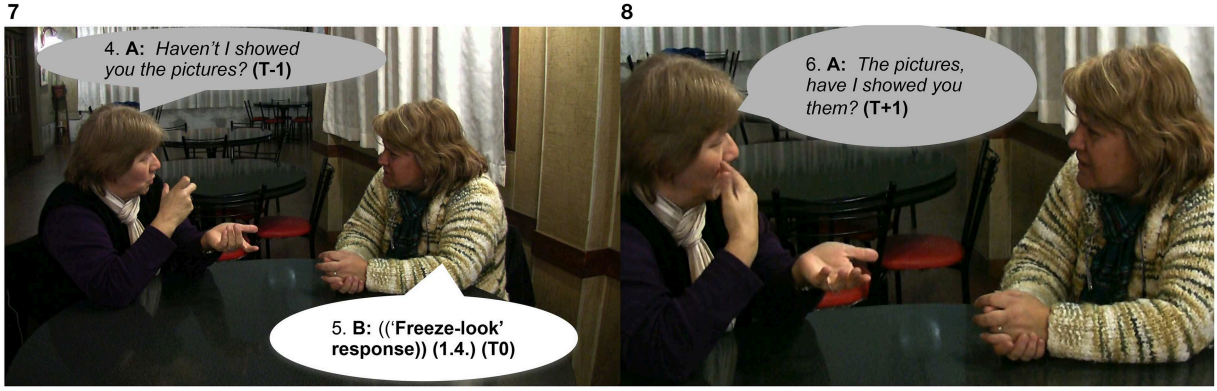


FIGURE 7 | *“Haven't I showed you the pictures?,”* Signer A asks a question to Signer B, sitting on the right. Signer B suspends her body position producing a “freeze-look.”

FIGURE 8 | *“The pictures, have I showed you them?,”* Signer A, sitting on the left, repeats the question and Signer B continues maintaining the “freeze-look” until toward the end of the repetition of the question, when she provides a fitted answer to A.

Summary: Extract 9

3	A	Do you know SN?	T−1
4	B	**((‘Freeze-look’ response))**	**T0**
5	A	Sign-name?	T+1

Extract 9. CH-LR | 105018

1 A HEY
Hey ((trying to get B’s attention))

2 B ((Stops signing movement, holding the last previous sign))

Q-ET
3 A KNOW SN-H (1.1) T−1
Do you know ((sign-name))?

4 B **((‘Freeze-look’ response))** (2.1) **T0**

Q-ET
5 A SN---H T+1
((Sign-name?))

6 B ((Continues in ‘freeze-look’ position))

7 B NO NO
No, no ((Drops freeze position to no signing position))

Extract 9 shows a similar case, beginning with an attentional problem: Signer B is not looking at Signer A, and Signer A has to secure B’s attention in order to proceed. Once Signer B’s visual attention is on Signer A, Signer A then asks B about another person using a sign name (SN), in line 3. From the beginning of A’s question in Line 3, Signer B produces a “freeze-look,” looking directly at Signer A and holding still his signing position. Then in Line 5, Signer A repeats the sign name of the person he has asked about. Note that in this case, B’s “freeze-look” is held for some time after the end of Signer A’s repeated question, and is released only when Signer B begins providing a fitted response.

The sequences in Extract 8 and Extract 9 illustrate the kinds of seeing problems that are common in sign language interaction, and they show that these problems can occur in the run-up to a “freeze-look” sequence. The two examples have a similar structure: Person A asks Person B a question, but B is not attending and fails to respond; Person A then secures B’s attention before repeating the question; Person B produces a “freeze-look” response; and finally Person A repeats the question and a fitted answer can be given, thus closing the sequence and allowing the conversation to move forward. These cases help us see a distinction between non-response due to absence of attention (not seeing that one had been asked a question at all) and the open signal of non-response that we term the “freeze-look.” The key difference is revealed in how the non-response is treated by person A. If the non-response is simply due to B’s lack of attention, then A will then secure the required attention in some way. If the non-response is in the form of a “freeze-look” from B, then B will repeat the question. Because B is looking directly at A when they produce a “freeze-look,” then the problem cannot be one of attention or perception; instead, because B is studiously not responding, the implication is that they cannot respond, and this will most likely be because they have not clearly comprehended what was just asked. The simple solution is for Signer A to repeat the question: precisely the response that they would have produced had Signer B asked “What did you say?.”

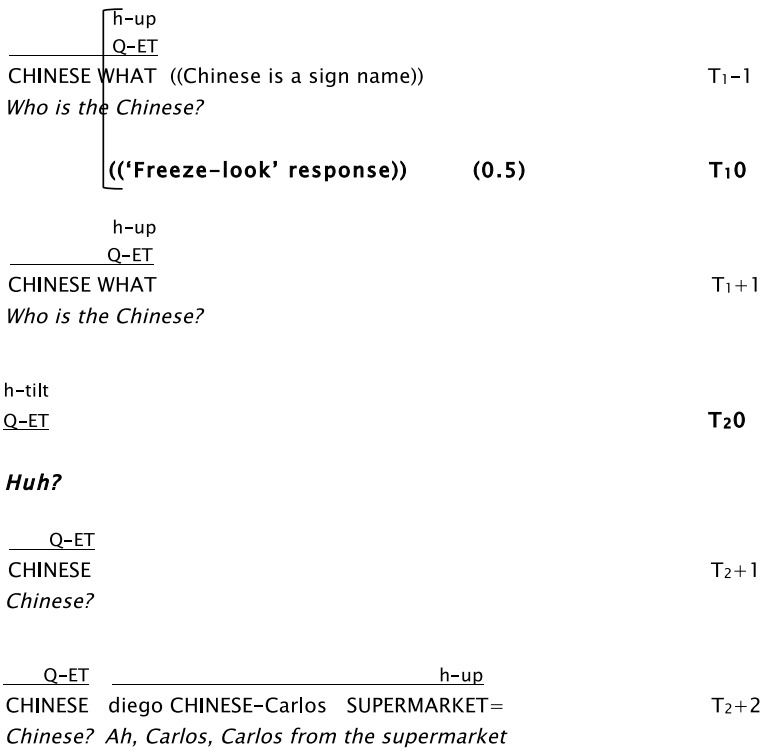
Pursuit Cases: From Implicit to Explicit OIR

If we are correct in claiming that the “freeze-look” is an *off-record* way to do other initiation of repair, this implies that it is at the “weak” end of the scale of repair strategies (Schegloff et al., 1977, p. 369). This would lead to the following prediction: If a “freeze-look” response to a question does not elicit a repetition or clarification of that question, the person who produced the “freeze-look” can then upgrade to a more explicit or on-record initiation of repair. This prediction is borne out in the LSA corpus. Almost 50% of the “freeze-look” action cases (11 out of 23) are upgraded to an explicit on-record OIR (while the opposite ordering is not observed). In most of the observed cases, a “freeze-look” is upgraded to an open format of other-initiation of repair (such as *What?*), but it may also be upgraded to a restricted format (such as *Who? Where?)*. We now look at some examples.

Summary: Extract 10

1	A	Who is the Chinese (person)?	T_1−1
2	B	**((‘Freeze-look’-response))**	**$T_1$0 ((off-record))**
3	A	Who is the Chinese?	T_1+1
4	B	**Huh?**	**$T_2$0 ((on-record))**
5	A	The Chinese?	T_2+1
6	B	Chinese? Ah, Diego, Diego from the supermarket.	T_2+2

Extract 10. SP | 381107



In Extract 10 Signer B’s first response (in Line 2) is a “freeze-look,” but the second version of the question produced by Signer A in response (in Line 3) does not appear to be adequate. Rather than giving an answer to the question, Signer B instead upgrades to an explicit way of initiating repair (i.e., a head tilt that can be translated as “Huh?”). The problem is eventually solved, with Signer B able to answer the question in line 6.

In Extract 11, Signers A and B are chatting about B’s son, who is working on a cruise. Signer A asks B if there are many tourists traveling on the cruise (line 1). After a “freeze-look” from Signer B, A repeats the question, with changes to the word order. At this point, rather than answering the question, in line 4 Signer B initiates repair more explicitly, with a sign that combines puckering of the lips and leaning forward of the head (also roughly translatable as “Huh?”), along with mouthing of the Spanish word *Como?*. Finally, in line 5, Signer A solves the sequence by partially repeating the question *“Tourism?,”* adding mouthing to the partial repetition.

In Extract 12, Signer A asks a question using only mouthing (in line 1), in response to which Signer B produces a “freeze-look” response as shown in **Figure 9**. Signer A does a repeat of the question, again using mouthing. Signer B’s response is now a more explicit type of OIR using several NMMs. He produces an open-mouth gesture (resembling the interjection *“huh?”),* raises his eyebrows and moves his head upwards (see **Figure 10**), while holding the manual signs produced in the previous turn (se) to explicitly initiate repair. Then, Signer A uses fingerspelling to solve the problem in line 5. Mouthing is a common cause of understanding problems in LSA that is often fixed by using fingerspelling instead (Manrique, in press).

Timing of “Freeze-looks”

Our study focuses on the function of the “freeze-look” in a specific context (immediately after a question) and identifies a specific function in that context (it elicits a “repair” of the question in the form of a repeat or near-repeat). In addition to measuring the effects of the “freeze-look” by examining the responses it elicits, we also measured aspects of the “freeze-look’s”

Summary: Extract 11

1	A	Are there many tourists?	T_1−1
2	B	**((‘Freeze-look’ response))**	**$T_1$0**
3	A	Are there too many tourists?	T_1+1
4	B	Huh?	**$T_2$0**
5	A	Tourism?	
6		Are there many tourists coming?	T_2+1
7	B	Many tourists coming	
8		Yes, a lot	T_2+2

Extract 11. MS | 362670

		h-down Q-ER		
1	A	VISIT TOURISM COME=		T_1−1
		Are there many tourists?		
2	B	**=((‘Freeze-look’ response))**	**(0.5)**	**$T_1$0**
		h-down Q-ER		
3	A	=COME COME TOURISM-	H ((Holding last sign-Tourism))	T_2+1
		Are there too many tourists?		
4	B		LPF-H-blink ((Holding head and lips)) Mth: como? *Huh?*	**$T_2$0**
		Q-ER		
5	A	=TOURISM-H tourism?((mouthing)) *Tourism?*		
6			TOURISM TOURISM, COME COME? *Are there many tourists coming?*	T_2+1
7	B	Tourism	tourism PU come come TOURISM TOURISM *Many tourists coming*	
8		YES A-LOT *Yes, a lot*		T_2+2

timing. A first measure to note here is the response latency, i.e., the time between the end of the trouble source by Person A (with B producing a “freeze-look”) and the beginning of their “repair” or (near-) repeat of the original question. See **Figure 11**.

This shows that the “freeze-look” has a rapid effect in interaction: when a person finishes their question and finds that they are faced by their addressee (still) producing a “freeze-look,” then they will quickly follow up with a repeat or near-repeat of the question.

A second timing measure to note is the absolute duration of “freeze-look” behaviors. In our LSA data, there is a range in duration from 0.3 to 6.3 s, with 69 percent between 0.5 and 3 s. See **Figure 12**.

This suggests that “freeze-looks” can continue well past the onset of the “repair” that they elicit, in T+1. This is indeed the case: They tend to be “released” not at the moment at which the subsequent repair turn begins, but rather at the moment at which Person B is ready to produce their next utterance. This will either be when Person B upgrades to a “stronger” OIR because the repair was inadequate for some reason (which happens about 50% of the time), or it will be when Person B produces an “uptake” or similar turn that signals subsequent

Summary: Extract 12

1	A	((A mouths a company's name))		T_1-1
2	B	**((‘Freeze-look’ response))**	**((Figure 9))**	**$T_1$0**
3	A	((A repeats mouthing of a company's name))		T_1+1
4	B	**Huh?**	**((Figure 10))**	**$T_2$0**
5	A	Ehh, Cele (...) ((use of fingerspelling))		T_2+1
6	B	Hey, ((company's name)) no, it's not open anymore		T_2+2

Extract 12. BT | 51450

1	A	Mth+((company's name))-H *((Company's name))*		T_1-1
2	B	**((‘Freeze-look’ response)) (2.3)**	**((Figure 9))**	**$T_1$0**
3	A	mth: ((company's name))-H *((Company's name))*		T_1+1
		h-up Q-ER		
4	B	((holding hands from previous turn)) **OM-A:** *Huh?*	**((Figure 10))**	**$T_2$0**
5	A	H- ((hesitation)) C-E+ L-E mth: (...) *Ehh, Cele (...)*		T_2+1
		h-up		
6	B	PT-HEY NO+ THERE-IS-NOT Mth: ((company's name)) *Hey, ((company's name)) no, it's not open anymore*		T_2+2

FIGURE 9 | *((“Company's name”))*, Signer A, the woman sitting on the left side of the table, asks a question to Signer B, the man in the white t-shirt sitting on the right side of the table. At the end of the question, Signer B produces a “freeze-look” (lines 1 and 2) for 2.3 s.

FIGURE 10 | *“Huh?,”* Signer B, after Signer A repeats the question, initiates a more explicit type of OIR using several NMMs. He produces an open-mouth gesture (resembling “huh?”), raises his eyebrows and moves his head upwards, while holding the manual signs produced in the previous turn (see **Figure 8**) to explicitly initiate repair.

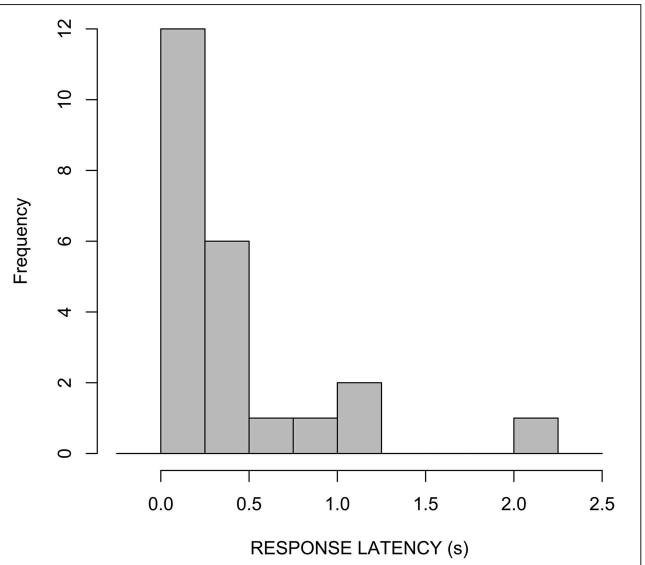


FIGURE 11 | Response latencies from onset of B's "freeze-look" to beginning of A's repair turn.

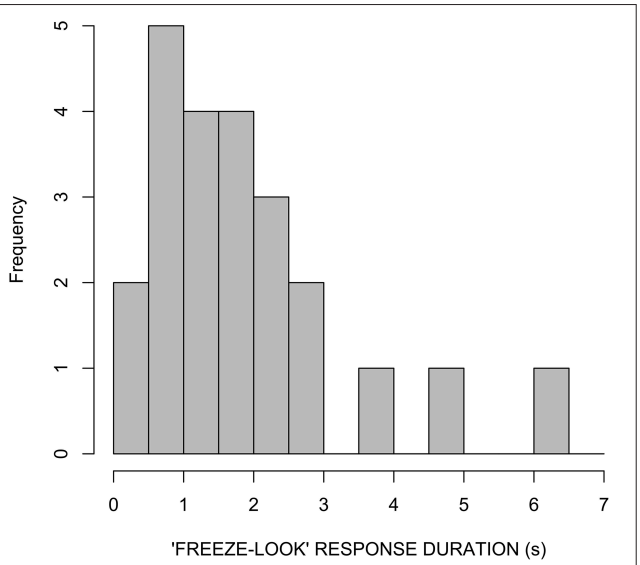


FIGURE 12 | "Freeze-look" durations in LSA data set.

repair turn was a satisfactorily resolution of the problem (see Floyd et al., 2014 for description of this timing pattern in a three-language comparison, involving LSA and two spoken languages; the form of the "freeze-look" and "hold" in spoken languages is similar to sign language, as it involves the same manual and facial articulators, being used for co-speech gesture).

Conclusion

The evidence we have presented from LSA shows that the "freeze-look" behavior—the act of keeping the whole body in a still position while looking directly at the person who has just asked a question—functions as an open-class other-initiator of repair (OIR), and additionally that it is "off-record" and somewhat weak in nature. Our claim that a "freeze-look" is a kind of OIR is supported by the fact that it gives rise to the same functional outcome as other known types of OIR: namely, it leads to a "re-doing" of the first utterance (e.g., a repeat or a reformulation). **Figure 13** summarizes the possible patterns of response and counter-response after a question that have been reviewed in this study, showing explicitly the functional identity or similarity of the "freeze-look" and other available OIR strategies.

Our claim that the "freeze-look" is an off-record strategy of OIR is firstly based on its intrinsic semiotic properties: it does not use symbolic resources (i.e., conventional signs such as lexical items that have semantic entailments) to explicitly encode that there is a problem of understanding; instead, it uses non-symbolic resources (i.e., indexical signals of the kind that occur in animal communication; on these semiotic distinctions see Enfield, 2013: Chapter 4). Secondly, its status as off-record is consistent with the fact that it appears to be "weaker" than other available OIR options (just as an indirect request is "weaker" than a direct request). One sense in which it is weak is that it only seems to succeed half of the time it is used: in 50% percent of cases, a "freeze-look" is followed up by a stronger other-initiation or repair. The common upgrading of a "freeze-look" to an explicit OIR (such as "What?") shows that it occupies a position in a "paradigm" of alternative types of OIR. This was shown in the cases presented in Section Freeze-look: A Notable Absence of Response, above. In each of those cases, Signer B initially produces a "freeze-look" response to Signer A's question, leading to a re-doing of that question; however, the solution appears to be inadequate, and Signer B then upgrades with a stronger, on-record open OIR, indicating that Signer A's first re-doing of the question did not solve the problem. This ordering of Person B's chosen strategies for OIR in these sequences provides evidence in favor of the argument that the "freeze-look" is a weak type, which sometimes needs to be upgraded or strengthened. We propose that this can be captured by placing the "freeze-look" at the extreme "weak" end of a continuum of types of conversational repair (from cf. Schegloff et al., 1977; Sidnell, 2010). Schegloff et al. (1977) rank the OIR formats in in terms of their "strength" in identifying the trouble source of the OIR sequence. "Open-class" repair initiators (Drew, 1997) have been placed on the "weakest" end, as they leave open the identification of the trouble source; often, the entire previous turn needs to be re-done by the signer/speaker of the trouble source. On the "strongest" end of the continuum are "understanding-check" formats such as repetition of part or all of a previous turn that invite confirmation that what one just heard or understood was correct. Our proposal for expansion of the continuum is illustrated in **Figure 14**.

We do not want to imply that the off-record OIR function is the only function of the general behavior of holding the body still while looking at one's interlocutor. The findings of our study apply exclusively to the function of this practice

	A: (T-1)	B: (T0)	A: (T+1)
1	Q	Gives answer	Continuation
2	Q	OIR	Repeat or (near-)repeat
3	Q	Non-attention	Secure attention
4	Q	Thinking face (word search)	Waits for answer and/or helps B
5	Q	'Freeze-look'	Repeat or (near-)repeat

FIGURE 13 | Possible patterns of response and counter-response after a question, showing functional similarities between "freeze-look" and on-record OIR strategies.

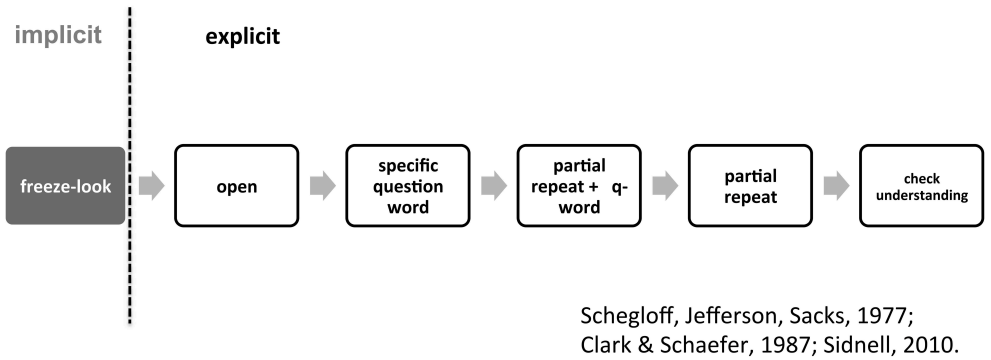


FIGURE 14 | Types of other initiation of repair laid out on a continuum from "weak" to "strong" (after Schegloff et al., 1977; Sidnell, 2010) with the "freeze-look" behavior placed at the extreme "weak" end of types of conversational repair.

in a specific position in a conversational sequence: i.e., just after a person has been asked a question. Further research is needed to investigate other functions that this behavior may have in other defined types of context, both in LSA, and cross-linguistically.

Finally, our data have come from a sign language, and so one might ask: Is this practice exclusive to sign language? It may not be surprising that we have noticed the "freeze-look" phenomenon in sign language conversation, given that visual behavior is obviously the exclusive focus of attention in this type of language. But users of spoken languages also have a rich set of visual resources at their disposal (McNeill, 1992; Kendon, 2004; Goldin-Meadow, 2005; Enfield, 2009). The "freeze-look" behavior can in principle be produced by anybody in a face-to-face setting, and so we may ask whether it is also used for other-initiation of repair in spoken languages. Only further research will tell, but we see no reason to think it would not be used in this way universally. At least this is a hypothesis to be tested. If the "freeze-look" turns out to be systematically used in spoken language interaction as well, then this study will have made a contribution not only to research on sign languages and on practices for other-initiation of repair in conversation, but it will have taken insights from research on sign language as pointers to an underexplored realm of possibility in spoken language: the systematic use of visible bodily behavior as part of the system of language.

Acknowledgments

This research was conducted as part of the European Research Council project *Human Sociality and Systems of Language of Use* (European Research Council Starting Grant 240853 HSSLU), directed by Nick Enfield and hosted by the Max Planck Society. We thank all the participants in *The Argentine Sign Language (LSA) Conversational Corpus in LSA (2010-2014)* and the *Association de Sordomudos de Ayuda Mutua* (ASAM); *Asociación Argentina para Sordomudas, Casa Hogar; Confederación Argentina de Sordos* (CAS); *Asociación Argentina de Sordos* (ASO); and *CRESCOMAS* San Juan, Argentina. We would especially like to thank Tony Druetta, as well as Juancito and Marita Druetta, Pablo Lemmo, Cecilia Maltez, Diego Morales, and Graciela Ruoti. We gratefully acknowledge the support of the Language and Cognition Department of the Max Planck Institute for Psycholinguistics, Nijmegen, directed by Steve Levinson, for providing an excellent institutional and intellectual environment. We are grateful to all participants in the Other-Initiated Repair research project at the MPI Nijmegen for comments and feedback: Julija Baranova, Joe Blythe, Penny Brown, Mark Dingemanse, Paul Drew, Tyko Dirksmeyer, Simeon Floyd, Rosa Gisladottir, Kobin Kendrick, Stephen Levinson, and Giovanni Rossi. For comments and feedback we also thank Onno Crasborn, Herb Clark, John Heritage, Galina Bolden, Adam Kendon, and Trevor Johnston. We would also like to express

our gratitude to two anonymous reviewers for providing critical comments on an earlier version of this manuscript and greatly helping to improve it. Finally, we would like to thank Mart Lubbers, Francisco Torreira, and Michael Wiemers for technical support with Python and R.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01326

References

Baker, C. (1977). "Regulators and turn-taking in American Sign Language discourse," in *On the Other Hand: New Perspectives on American Sign Language,* ed L. Friedman (New York, NY: Academic Press), 215–236.

Baker, C., and Padden, C. (1978). "Focusing on the nonmanual components of American Sign Language," in *Understanding Language Through Sign Language Research,* ed P. Siple (New York, NY: Academic Press), 27–57.

Baker-Shenk, C. L. (1983). *A Microanalysis of the Nonmanual Components of Questions in American Sign Language.* Doctoral dissertation, University of California, Berkeley.

Behares, L., Massone, M., and Curiel, M. (1990). *El Discurso Pedagógico de la Educación del Sordo. Construcciones de Saber y Relaciones de Poder*. Buenos Aires: Facultad de Filosofía y Letras; Universidad de Buenos Aires.

Boyes-Braem, P., Sutton-Spence, R., and te Leiden, R. (eds.). (2001). *The Hands Are the Head of the Mouth: The Mouth as Articulator in Sign Languages*. Hanburg: Signum Press.

Brown, P., and Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge, UK: Cambridge University Press.

Clark, H. H. (1996). *Using Language.* Vol. 4. Cambridge, UK: Cambridge University Press.

Clark, H. H., and Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition* 84, 73–111. doi: 10.1016/S0010-0277(02)00017-3

Curiel, M., and Massone, M. I. (2004). Sign order in argentine sign language. *Sign Lang. Stud.* 5, 63–93. doi: 10.1353/sls.2004.0023

de Vos, C., Torreira, F., and Levinson, S. C. (2015). Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. *Front. Psychol.* 6:268. doi: 10.3389/fpsyg.2015.00268

Dingemanse, M., and Enfield, N. J. (2015). Other-initiated repair across languages: towards a typology of conversational structures. *Open Linguist.* 1, 98–118. doi: 10.2478/opli-2014-0007

Dingemanse, M., Kendrick, K., and Enfield, N. J. (in press). A coding scheme for other-initiated repair across languages. *Open Linguist.*

Dingemanse, M., Roberts, S., Baranova, J., Blythe, J., Drew, P., Floyd, S., et al. (in press). Universal principles in the repair of communication problems. *PLoS ONE.*

Dingemanse, M., Torreira, F., and Enfield, N. J. (2013). Is "Huh?" a Universal Word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS ONE* 8:e78273. doi: 10.1371/journal.pone.0078273

Dively, V. L. (1998). "Conversational repair in ASL," in *Pinky Extension and Eye Gaze: Language Use in Deaf Communities,* ed C. Lucas (Washington, DC: Gallaudet University Press), 137–169.

Drew, P. (1997). "Open" class repair initiators in response to sequential sources of troubles in conversation. *J. Pragmat.* 28, 69–101. doi: 10.1016/S0378-2166(97)89759-7

Emmorey, K. (2001). *Language, Cognition, and the Brain: Insights From Sign Language Research*. Mahwah, NJ: Psychology Press.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances, Vol. 8*. Edited by S. C. Levinson. Cambridge: Cambridge University Press.

Enfield, N. J. (2013). *Relationship Thinking: Agency, Enchrony, and Human Sociality*. New York, NY: Oxford University Press.

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., Dirksmeyer, T., et al. (2013). "Huh? What?—A first survey in 21 languages," in *Conversational Repair and Human Understanding,* eds M. Hayashi, G. Raymond, and J. Sidnell (Cambridge: Cambridge University Press), 343–380.

Floyd, S., Manrique, E., Rossi, G., and Torreira, F. (2014). The timing of visual bodily behavior in repair sequences: evidence from three languages. *Dis. Proces.* 52, 1–30. doi: 10.1080/0163853x.2014.992680

Goldin-Meadow, S. (2005). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Harvard University Press.

Goodwin, C., and Goodwin, M. H. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica* 62, 51–75.

Hayashi, M., Raymond, G., and Sidnell, J. (eds.). (2013). *Conversational Repair and Human Understanding*. Cambridge: Cambridge University Press.

Jefferson, G. (2004). "Glossary of transcript symbols with an introduction," in *Conversation Analysis: Studies from the First Generation* (Amsterdam: John Benjamins Publishing), 13–34.

Jefferson, G. (2015). *Talking About Troubles in Conversation*. New York, NY; Oxford: Oxford University Press.

Johnston, T. (2010). From archive to corpus: transcription and annotation in the creation of signed language corpora. *Intern. J. Corpus Linguist.* 15, 106–31. doi: 10.1075/ijcl.15.1.05joh

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge, UK: Cambridge University Press.

Klima, E. S., and Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition* 14, 41–104. doi: 10.1016/0010-0277(83)90026-4

Levinson, S. C., and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 6:731. doi: 10.3389/fpsyg.2015.00731

Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge, UK: Cambridge University Press.

Manrique, E. (2011). *Other-Repair Initiators in Argentinean Sign Language: Handling Seeing and Understanding Difficulties in Face-to-face Interaction*. Nijmegen: Radboud Universiteit.

Manrique, E. (in press). Other-initiated repair in argentine sign language. *Open Linguist.*

Massone, M. I. (1993). *Lengua de Señas Argentina: Diccionario Bilingue = Argentine Sign Language: Bilingual Dictionary, Vol. 1–2*. Buenos Aires: Sopena.

Massone, M. I., and Curiel, M. (1993). Categorías gramaticales en la lengua de señas argentina. *Rev. Lingüíst. Teóric. Apl.* 31, 27–54.

Massone, M. I., and Machado, E. M. (1992). *Lengua de Señas Argentina: Análisis y Vocabulario Bilingüe*. Buenos Aires: Edicial.

McCleary, L., and Leite, T. de A. (2013). Turn-taking in Brazilian sign language: evidence from overlap. *J. Interact. Res. Commun. Disord.* 4:123. doi: 10.1558/jircd.v4i1.123

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.

Mesch, J. (2001). *Tactile Sign Language: Turn Taking and Question in Signed Conversations of Deaf-Blind People*. Washington, DC: Gallaudet University Press.

Pfau, R., and Quer, J. (2010). "Nonmanuals: their grammatical and prosodic roles," in *Sign Languages,* ed D. Brentari (Cambridge: Cambridge University Press), 381–402.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735.

Sandler, W., Meir, I., Dachkovsky, S., Padden, C., and Aronoff, M. (2011). The emergence of complexity in prosody and syntax. *Lingua* 121, 2014–2033. doi: 10.1016/j.lingua.2011.05.007

Sandler, W., and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge, UK: Cambridge University Press.

Schegloff, E. A. (1968). Sequencing in conversational openings. *Am. Anthropol.* 70, 1075–1095. doi: 10.1525/aa.1968.70.6.02a00030

Schegloff, E. A. (1982). "Discourse as an interactional achievement: some uses of "Uh Huh" and other things that come between sentences," in *Analyzing Discourse: Text and Talk,* ed D. Tannen (Washington, DC: Georgetown University Press), 71–93.

Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language* 53, 361–382. doi: 10.1353/lan.1977.0041

Schegloff, E. A., and Sacks, H. (1973). Opening up closings. *Semiotica* 8, 289–327. doi: 10.1515/semi.1973.8.4.289

Sidnell, J. (2010). *Conversation Analysis: An Introduction*. London: Wiley-Blackwell.

Sifianou, M. (1999). *Politeness Phenomena in England and Greece: A Cross-Cultural Perspective*. Oxford: Oxford University Press.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. USA* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Stivers, T., and Robinson, J. D. (2006). A preference for progressivity in interaction. *Lang. Soc.* 35, 367–392. doi: 10.1017/S0047404506060179

Sutton-Spence, R., and Woll, B. (1999). *The Linguistics of British Sign Language: An Introduction*. Cambridge: Cambridge University Press.

Valassina, S. (ed.). (1997). *Diccionario Lengua de Señas Argentina—Español*. Buenos Aires: Ministerio de Culture y Educación de la Nación.

Veinberg, S. (1993). La función linguística de los rasgos no-manuales en la aseveración, la negación y la interrogación en la Lengua de Señas Argentina. *Signo Seña* 2, 101.

Veinberg, S. (1996). Argentinien: Anfänge und Entwicklung der Erziehung Gehörloser (Argentina: Inicios y desarrollo de la educación del sordo). *Das Zeichen. Zeitschrift Zum Thema Gebärdensprache Und Kommunikation Gehörlose* 38, 488–496.

Vermeerbergen, M., Leeson, L., and Crasborn, O. A. (2007). *Simultaneity in Signed Languages: Form and Function*. Amsterdam: John Benjamins Publishing.

Willoughby, L., Manns, H., Shimako, I., and Bartlett, M. (2014). Misunderstanding and repair in tactile auslan. *Sign Lang. Stud.* 14, 419–443. doi: 10.1353/sls.2014.0014

Wittenburg, P., et al. (2006). "Elan: a professional framework for multimodality research," in *Proceedings of Language Resources and Evaluation Conference (LREC)*. Available online at: http://tla.mpi.nl/tools/tla-tools/elan/

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 September 2015
doi: 10.3389/fpsyg.2015.01167

Early development of turn-taking in vocal interaction between mothers and infants

Maya Gratier[1], Emmanuel Devouche[2], Bahia Guellai[1], Rubia Infanti[1,3], Ebru Yilmaz[1] and Erika Parlato-Oliveira[3]*

[1] Université Paris Ouest Nanterre La Défense, Paris, France, [2] Université Paris Descartes, Paris, France, [3] Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Infants are known to engage in conversation-like exchanges from the end of the second month after birth. These 'protoconversations' involve both turn-taking and overlapping vocalization. Previous research has shown that the temporal organization of adult–infant turn-taking sequences is similar to that of adult verbal conversation. It has also been shown that young infants adjust the quality of their vocalization in response to the quality and timing of adult vocalization. We present new evidence of turn-taking interaction in infants aged between 8 and 21 weeks based on the analysis of 176 samples of naturalistic face-to-face interactions from 51 dyads. We found high levels of latched turns as well as frequent initiation of turn-taking by infants at these ages. Our data do not support the hypothesis that turn-taking ability increases with age between 2 and 5 months but do suggest that infants are active participants in turn-taking from the earliest age and that mothers adjust turn-taking formats to infants.

Keywords: early mother–infant communication, turn-taking, protoconversation, vocal development

OPEN ACCESS

Edited by:
Marisa Casillas, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Elma Hilbrink, Max Planck Institute for Psycholinguistics, Netherlands
Ann Bigelow, St. Francis Xavier University, Canada

****Correspondence:***
Maya Gratier, Laboratoire Ethology, Cognition, Développement (BSL114), Université Paris Ouest Nanterre La Défense, 200 Avenue de la République, 92001 Nanterre, Paris, France
gratier@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *06 February 2015*
Accepted: *24 July 2015*
Published: *04 September 2015*

Citation:
Gratier M, Devouche E, Guellai B, Infanti R, Yilmaz E and Parlato-Oliveira E (2015) Early development of turn-taking in vocal interaction between mothers and infants. Front. Psychol. 6:1167. doi: 10.3389/fpsyg.2015.01167

Introduction

Conversation is a complex communicative process in which visual and auditory signals are combined in a flow of turn taking. An important aspect of the temporal organization of conversation is the fluency between turns. The end of a speaker's turn needs to be anticipated by the next speaker for the conversation to be fluent (Sacks et al., 1974). In everyday conversation, people are remarkably rapid and accurate in switching between listener and speaker roles (Magyari and De Ruiter, 2012). The gap between turns in conversation between native speakers of the same language is generally around 250 ms (De Ruiter et al., 2006; Stivers et al., 2009). Speakers predict not only when a turn will end but also the content of the turn, enabling them to respond in a well-timed and semantically appropriate manner. Computational cognitive models fail to account for the rapidity and efficiency of transitions between speakers because they have ignored the role of embodiment in predicting speaker turns (Goodwin, 1986). Experimental studies suggest that speakers rely on both lexico-syntactic information and pragmatic signals to project the end of a turn but that accurate predictions are not possible without lexical content (De Ruiter et al., 2006).

Yet a number of studies in psychology describe turn-taking interaction occurring between infants and adults as early as in the second month of life, long before any access to lexical information. A recent study conducted by Caskey et al. (2011) on premature infants showed that as

Abbreviations: CI, confidence interval; IM, infant–mother; MI, mother–infant; OR, odds ratio; TTS, turn-taking sequence.

early as 32–36 weeks gestational age the frequency of infant vocalization increases in the presence of a parent. Preterm infants were also found to produce reciprocal vocalizations, supporting the hypothesis that turn-taking is a precocious human ability (Caskey et al., 2011). Stivers et al. (2009) have shown the powerful cross-cultural stability of two basic rules of the turn-taking system, avoiding overlap and minimizing silence between turns. Do young infants already know these rules or do they learn them in the course of interacting with partners? Few studies offer a developmental account of turn-taking interaction in young infants, prior to language ability.

In the first weeks after birth, infants begin to produce vocalizations that are described in the literature as coos and murmurs (Oller, 2000) and that elicit emotional and motivated responses from social partners. These vocalizations are not readily associated with either positive or negative emotion, in the way that laughter or crying are (Oller et al., 2013). They are produced when infants are alert and relaxed or playful. They are frequently associated with an intent knit-brow gaze directed at the social partner and with mouth movements resembling those of speech that have been called pre-speech movements (Trevarthen, 1993). Both caregivers and naive observers interpret these vocalizations as being produced intentionally, wilfully and with effort (Bloom and Lo, 1990; Beaumont and Bloom, 1993).

The appearance of these kinds of vocalizations gives rise to vocal exchanges with adult partners that resemble conversations because they are characterized by alternating vocalizations separated by audible pauses (Bateson, 1975; Stern et al., 1975; Trevarthen, 1979; Bloom, 1988). Adults respond to the coos and murmurs of young infants with specific prosody and timing. They have been shown to closely match the acoustic qualities of infant vocalizations, producing short repetitive bouts of infant-directed speech aimed at eliciting further vocalization on the part of the infant (Papoušek and Papoušek, 1989; Gratier and Devouche, 2011). Mothers also modulate their pitch according to the perceived emotional expressions of infants (Smith and Trainor, 2008). It has also been demonstrated that infants in turn adjust the quality of their vocalization in response to adults. The kinds of vocalizations involved in vocal turn-taking with an adult have been described as having a 'speech-like' quality compared with vocalizations produced alone or outside of a turn-taking format (Bloom et al., 1987). Vocal interaction in the first months of life thus appears to be bidirectional and mutually regulated (Lavelli and Fogel, 2013).

Various approaches have been used to demonstrate the active role of both mothers and infants in the first half year of life in face-to-face interaction. A durational approach reveals stable timing structures in maternal behavior across vocal and kinesic modalities (Stern et al., 1977; Cossette et al., 1986). Mothers' speech to young infants is timed in such a way as to leave room for infant response. Maternal utterances are brief (between 0.5 and 1.5 s) and followed by pauses of around 1 s, usually followed by another utterance. Pauses between utterances that are connected through repetition of form, content, or topic rarely exceed 3 s and pauses longer than 3 s generally demarcate episodes of mutual engagement (Stern et al., 1977; Stern and Gibbon, 1979). With infants aged between 2 and 4 months, the majority of maternal responses to infants occur with a 1 s latency after the signal (Keller et al., 1999) and vocalization is responded to more frequently than gaze and smiling (Van Egeren et al., 2001). Pauses between alternating vocal turns of young infants and mothers have been found to range from 500 ms to 1 s (Jaffe et al., 2001).

A related approach to studying protoconversational organization focuses on the match or cross-correlation between temporal patterns of maternal and infant vocalization in spontaneous face-to-face interaction. The fact that mothers and infants engaged in social interaction match each other for vocalization and pause duration suggests a mutual regulation of the turn-taking exchange (Beebe et al., 1985; Jaffe et al., 2001). In her pioneering study of the protoconversations of a single infant aged between 6 and 13 weeks, Bateson (1975) highlighted the bidirectionality of infant and mother vocalization by showing that the mean duration separating successive utterances is longer when the previous utterance is by self that when it is by other. Both infant and mother respond to each other faster than they repeat an utterance of their own. Despite evidence that infants are capable of selective vocalization from the age of 2 months (Delack and Fowlow, 1978; Bloom, 1990; D'Odorico and Franco, 1991), there remains some controversy over the extent to which young infants actively contribute to turn-taking exchanges and the extent to which adults construct conversational frameworks for infant vocalization.

If infants partake in a truly co-regulated turn-taking, they must have the ability to perceive the contingent relations between their own behavior and that of their partner. These contingent relations hinge on the perception of timing in social interaction. Research has shown that infants' sensitivity to contingency changes right around the time when protoconversations appear, around the age of 2 months (Striano et al., 2005). The still-face paradigm has demonstrated that by 2 months infants react to the sudden interruption of social interaction by the mother with reduced smiling and gazing and increased fussiness and self-comforting behavior (Tronick et al., 1978). Murray and Trevarthen's (1985, 1986) closed-circuit double-TV paradigm shows that infants as young as 6 weeks are sensitive to social contingencies. When infants and mothers are made to interact via a live closed-circuit television set up, they are able to establish mutual gaze and partake in relaxed protoconversational interaction involving imitation across multiple modalities. However, when a sequence from the previous recording is replayed to either mother or infant with a few minutes delay, each partner becomes perturbed, expressing anxiety and aloofness. A 3 s window appears to reflect contingency for a wide range of behaviors involved in spontaneous social interaction (Van Egeren et al., 2001). The timing of an infant's response to the partner has often been taken as a demonstration of the kind of active interpersonal coordination that underlies communication between adults.

Starting around the age of 2 months, the cooing stage (Oller, 1980; Stark, 1980) is associated with a marked rise in face-to-face interaction described as primary intersubjectivity (Trevarthen, 1977; Lavelli and Fogel, 2013). By the age of 4 months, however, infants' interest shifts from an intense involvement with other people to involvement with objects (Trevarthen and Hubley,

1978; Cohn and Tronick, 1987). In interactive object play, adults and infants become engaged in doing things together, such as exploring objects or using them in novel ways. It is not clear from the existing literature whether the shift from primary intersubjectivity to object play involves a change in the turn-taking organization of social interaction. Indeed, very few studies have focused on turn-taking in social interactions with infants involving object play or joint attention.

One longitudinal study has reported a quantitative increase in vocal turn-taking with less overlapping vocalization between 12 and 18 weeks of age (Ginsburg and Kilbourne, 1988) suggesting infants' turn-taking competence increases around 4 months of age. Another study supporting the hypothesis that infants' turn-taking competence increases with age was conducted by Rutter and Durkin (1987). Using both a transversal and a longitudinal design, this study focused on vocal and gaze coordination from the end of the first year to the end of the third year of life. According to these researchers, infants younger than 24 months more frequently interrupt their mothers, whereas after this age they begin to truly coordinate their vocalizations with those of the mother. They also found that infants use gaze to actively signal the end of a turn by the age of 18 months. At this age, their gaze patterns begin to resemble those used in adult conversation in that they indicate giving the floor to their interlocutor and confirm when the floor is about to be offered to them by looking up at the interlocutor at the end of a turn.

The most detailed longitudinal study of the timing of turn-taking in infancy was conducted recently by Hilbrink et al. (submitted) on 12 infants aged between 3 and 18 months. These researchers report on the prevalence of turn-taking exchanges throughout the period they studied. They also find that turn-taking organization varies little between 3, 4, and 5 months of age but that turns slow down markedly around the age of 9 months. At this age infants respond to mothers' utterances with longer gaps whereas maternal turn-timing remains stable over time. Furthermore, this study shows that the amount of overlap in turn-taking remains constant between 3 and 5 months of age but decreases thereafter. The slowing down of turn-taking can be associated with important qualitative changes in social interaction and communicative skills such as joint attention.

The present paper assesses developmental change in turn-taking organization between the 3rd and the 5th months of life. We first aimed to explore the hypothesis that turn-taking is, from the first weeks of life, a mutually coordinated effort and then to assess developmental change and stability in the amount of overlap, duration of switching pauses (gap) between turns and length of TTSs. We thus compared spontaneous vocal interactions between mothers and infants ages between 2 and 3 months with those of mothers with 4-to-5-months-old infants.

Materials and Methods

Participants

Fifty one MI dyads participated in this study. Infants were aged between 8 and 21 weeks [28 boys and 23 girls aged respectively 12.8 weeks ± 3.77 (range: 8–19) and 13.2 weeks ± 3.73 (range: 8–21)], were born full-term and in good health. Out of the 51 mothers, three mothers were bilingual speakers but spoke to their infants in French. The sample was divided into two age groups: a group of 'younger' infants (35 8–13 weeks, mean age 10.8 weeks ± 1.54, 15 girls) and a group of 'older' infants (15 17–21 weeks, mean age 18.3 ± 1.24, 8 girls). The present research was approved by the university ethics committee (CCP n°1450089).

Apparatus and Procedure

All dyads were recorded in naturalistic contexts, in their home, when infants were in a quiet alert state. Before each observation, consent forms were signed by the parents. Mothers and infants were placed in comfortable positions, facing each other. Mothers were asked to talk to their infants in their usual manner for approximately 10 min and to avoid using any toys. Video and audio recordings were made using two camcorders (Sony Handycam HDR-CX190) and a digital audio recorder (Korg Sound on Sound Unlimited Track Recorder) placed near the dyad. Only the audio recordings were used in this study.

Acoustic Analysis

Selection of the Audio Samples

A total of 176 audio samples were selected [3.5 audio samples on average per infant ± 2.2 (range: 1–9). Audio sample length was on average 106.7 s ± 57.9 (range: 18–252)]. In all 90 min and 41 s of interaction were analyzed. Samples lasted on average 31 s. The samples were taken from a large corpus of audio recordings, they were the first to meet following four selection criteria:

(1) Each sample included at least one vocal contribution by the infant and no negative infant vocalizations (fuss, cry).
(2) In each sample, mothers addressed their infant directly. Sequences including song to infants were discarded.
(3) Samples were segmented based on pause duration: pauses between vocalizations that exceeded 3 s marked the end of the selected audio sample.
(4) Recording quality was optimal for all samples.

Software

We used Sound Analysis Pro (Tchernichovski et al., 2000) to segment the sequences and to obtain acoustic measures of individual vocalizations. Sound Analysis Pro is used primarily in studies of birdsong but has recently been used successfully to study infant vocalization (Lipkind et al., 2013). Data were exported and manipulated in Excel.

Coding of Vocalizations and Pauses

Based on visualization of spectrograms and audio guidance, each sequence was manually segmented into 4 types of events: maternal vocalization, infant vocalization, overlapping vocalization and pause, according to the following criteria:

(1) A vocalization (either by mother or infant) was defined as the production of vocal sound by one partner that was either continuous or included unvoiced segments of less than 300 ms. If the silent pause following an audible vocal sound

was greater than 300 ms, two successive vocalizations were coded.

(2) Overlapping vocalization was coded when either mother or infant vocalized over the vocalization of the partner. The entire vocalization was coded as overlapping even if it was only partially masked by the partner's vocalization.

(3) Pauses occurred either between two vocalizations by the same partner (within- speaker pause) or between alternating vocalizations (switching pause). Within-speaker pauses had a duration that was necessarily greater than 300 ms and lower than 3000 ms. Switching pauses could range from a few milliseconds to 3000 ms. Two alternating vocalizations were qualified as latched when the switching pause had a duration of less than 50 ms.

(4) Vegetative sounds produced by infants such as burps, growls or hiccups, noise from the environment and vegetative sounds produced by mothers, such as coughs, were not coded.

Coding of Turn-Taking Sequences

A TTS was defined as sequence of vocalizations involving at least one alternation between speakers. Such a sequence could involve alternation between a vocalization of mother and of infant or between a vocalization of mother or infant and an overlapping vocalization. A TTS ended when the same speaker produced at least two vocalizations in succession or the pause following a vocalization was greater than 3000 ms.

Acoustic Measures

Sound Analysis Pro (Tchernichovski, 2012) provides automated analysis of various acoustic features of the segmented sounds. These include duration, amplitude, pitch, frequency modulation, and entropy. For this study only durations and frequencies of vocalizations and pauses were used.

Inter-Coder Reliability

Twenty percent of the data set were double-coded. Inter-coder reliability (Pearson product-moment correlations) was 0.87, 0.98, 0.78, and 0.95 respectively for number of infants' vocalizations, mothers' vocalizations, vocalizations involving overlap and pauses. Onset positions were considered identical if they occurred within 50 ms, thus measures of vocalization duration had an error tolerance of up to 100 ms. Both coders correctly identified 80.6% of all onset positions within the subset of double-coded sequences.

Statistics

Analysis was done with Stata for Windows (version 12). Multiple regression was used to compute partial regression coefficients and logistic regression to estimate ORs and Wald 95% CIs (Cohen et al., 2003). Age (8–13 weeks vs. older than 13 weeks) and gender were treated as binary variables and included in the model as factors, and dyads as potential confounders. A general linear model (GLM) was used to analyze number of vocalizations per minute, with infants' gender and age as factors and including partial eta square as index of effect size. A chi square test was applied on the contingency tables, including Cramer's V as an index of effect size. An alpha level of 0.05 was used for all statistical tests.

Results

Vocal Production

The entire sample comprised a total of 2943 vocalizations of which 748 were produced by the infants alone, 1851 were produced by the mothers alone and 344 involved both mothers and infants vocalizing in overlap. Thus, the total sample of infant vocalizations comprised 1092 vocalizations and the total sample of maternal vocalizations comprises 2195 vocalizations. In total 2152 pauses were identified, of which 838 were switching pauses (38.9%).

Infant Vocalizations

Infant vocalizations occurring within less than 3 s of a maternal vocalization, i.e., involved in turn taking, represented 73.1% of all infant vocalizations (see **Table 1**). Infant vocalizations which were not involved in turn taking were produced either after an infant vocalization (12.8%) or were isolated, i.e., were neither preceded nor followed by another vocalization (14.1%). The proportion of infant vocalization occurring within TTSs was higher among younger infants than among older infants (78.7% vs. 62.2%; $OR = 2.25$, 95% CI: 1.69 2.99, $p < 0.0001$).

Among the vocalizations occurring within TTSs, some were overlapping vocalizations ($n = 344$; 31.5%), and the remaining were vocalizations that followed a maternal vocalization ($n = 454$; 41.6%). These were either latched vocalizations, that is vocalizations occurring without an intervening pause ($n = 202$; 18.5%) or vocalizations involving a switching pause ($n = 252$; 23.1%; see **Table 1**). Contingency analysis conducted on type of vocalization revealed a significant age effect (Chi square = 39.8; $p < 0.0001$; Cramer's $V = 0.21$): older infants vocalized twice in succession more frequently than younger infants whereas vocalizations occurring after a switching pause were more frequent among the younger infants. No difference was observed between the two groups for frequency of overlapping vocalizations (see **Figure 1**).

Infant vocalizations lasted on average 869.7 ms ± 662.3 (range: 50.2–3640.0). Multiple regression conducted on the durations of vocalizations revealed no effect of age ($p = 0.49$), but a significant gender effect: girls' vocalizations were on average 108 ms longer than boys' vocalizations ($p = 0.049$).

Maternal Vocalizations

Maternal vocalizations occurring within less than 3 s of an infant vocalization – i.e., involved in turn taking – represented 35.8% of all maternal vocalizations (see **Table 2**). One in five mother vocalizations was consecutive to an infant vocalization (441; 20%), either as a latched turn ($n = 184$; 8%) or involving a switching pause ($n = 257$; 12%). Only 13.5% of mother vocalizations were not embedded in a sequence of vocalizations ($n = 297$) and 51% were preceded by a maternal vocalization ($n = 1113$).

Contingency analysis conducted on type of vocalization revealed a significant age effect (Chi square = 51.6; $p < 0.0001$;

TABLE 1 | Proportions of infant vocalizations according to position in relation to other vocalizations and pauses.

	Infant vocalizations (*N* = 1092)					
	All infants (*N* = 1092)		8–13 weeks (*N* = 719)		17–21 weeks (*N* = 373)	
	N	%	*N*	%	*N*	%
Vocalizations preceded by another infant vocalization	140	12.8%	66	9.2%	74	19.8%
Vocalizations involved in turn taking						
Preceded by mother vocalization with pause	252	23.1%	195	27.1%	57	15.3%
Preceded by mother vocalization without pause	202	18.5%	145	20.2%	57	15.3%
Involving overlap	344	31.5%	226	31.4%	118	31.6%
Isolated vocalizations						
Neither preceded nor followed by any other vocalization	154	14.1%	87	12.1%	67	18.0%

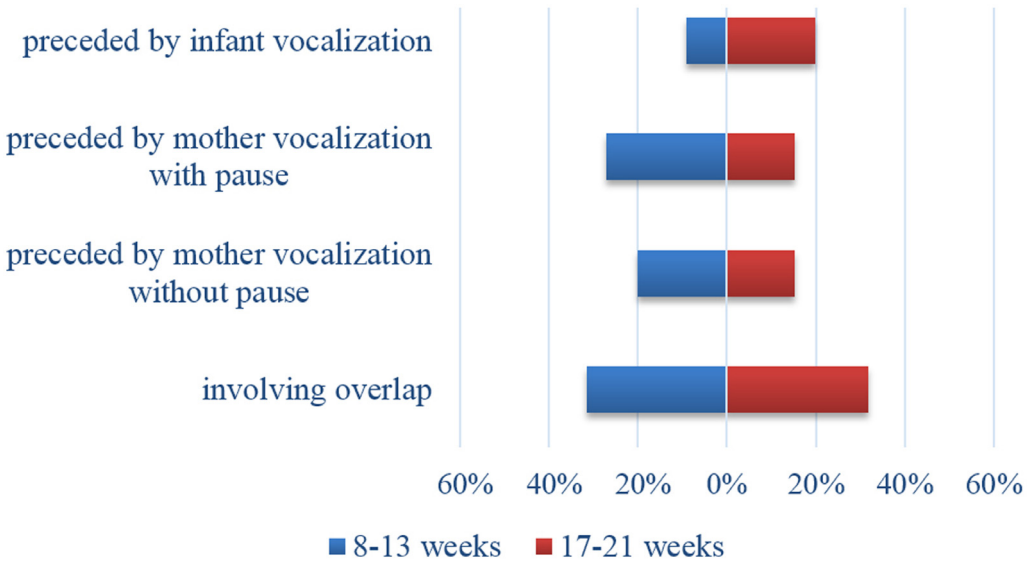


FIGURE 1 | Proportions of infant vocalization according to age.

TABLE 2 | Proportions of maternal vocalizations according to position in relation to other vocalizations and pauses.

	Mother vocalizations (*N* = 2195)					
	All mothers (*N* = 2195)		8–13 weeks (*N* = 1704)		17–21 weeks (*N* = 491)	
	N	%	*N*	%	*N*	%
Vocalizations preceded by another mother vocalization	1113	50.7%	928	54.5%	185	37.7%
Vocalizations involved in turn taking						
Preceded by infant vocalization with pause	257	11.7%	202	11.9%	55	11.2%
Preceded by infant vocalization without pause	184	8.4%	136	8.0%	48	9.8%
Involving overlap	344	15.7%	226	13.3%	118	24.0%
Isolated vocalizations						
Neither preceded nor followed by any other vocalization	297	13.5%	212	12.4%	85	17.3%

Cramer's $V = 0.16$). Mothers vocalized twice in succession more frequently with younger infants but overlap was more frequent for mothers of older infants (see **Figure 2**).

Mothers' vocalizations lasted on average 1144.7 ms ± 904.9 (range: 50–9803). Multiple regression conducted on the durations of the vocalizations showed a significant effect of age: the vocalizations of mothers of younger infants were on average 399 ms longer than those of mothers of older infants ($p < 0.0001$). A gender effect was also observed, mothers of boys vocalizing on average longer ($p = 0.034$).

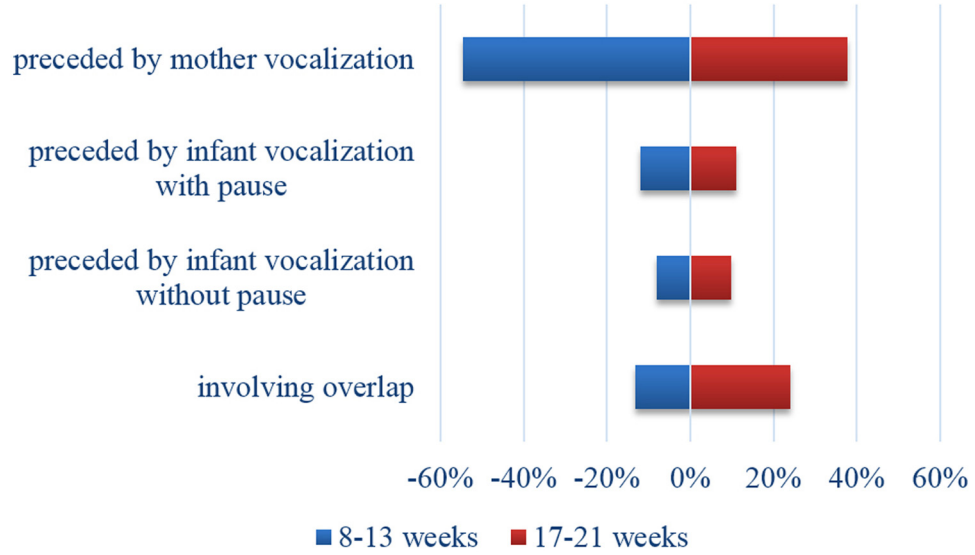


FIGURE 2 | Proportions of maternal vocalization according to infant age.

Vocalizations Involving Overlap

Vocalizations involving overlap lasted on average 1550.7 ms ± 1090.1 (range: 110–4928). Multiple regression conducted on the durations of the 344 overlapping vocalizations revealed a significant effect of age: overlapping vocalizations involving younger infants were 812 ms longer than those involving older infants ($p < 0.0001$). No gender effect was found ($p = 0.25$). Analysis of the number of vocalizations per minute revealed no significant effects, neither for gender ($p = 0.96$) nor for age ($p = 0.92$).

Latched Turns

Among the 454 turns produced by infants 44.5% ($n = 202$) were latched, i.e., without a pause between alternating vocalizations. This proportion was quantitatively higher for older infants than for younger ones but not significantly so (50.0% vs. 42.6%; $OR = 1.34$, 95% CI: 0.86 2.10, $p = 0.19$). Maternal vocalizations preceding a latched turn by the infant were shorter, though not significantly so, than those involved in turns including a switching pause (1089 ms vs. 1152 ms; $p = 0.35$). Among the 441 turns performed by mothers, 41.7% were latched. This proportion was quantitatively higher when mothers responded to older infants than when they responded to younger infants, but again not significantly so (46.6% vs. 40.2%; $OR = 1.29$, 95% CI: 0.81 2.07, $p = 0.26$). Infant vocalizations preceding a latched turn by the mother had the same duration as those involved in turns including a switching pause (respectively 869 and 870 ms).

Pauses

Within-Speaker Pauses

Pauses between vocalizations by the same speaker, which were by definition restricted to the range of 300 to 3000 ms, lasted on average 745.5 ms ± 557.0 (range: 310–2643.6) for infants and 967.3 ms ± 606.4 (range: 300–2994.8) for mothers. Multiple regression analysis revealed that within speaker pauses were not significantly different according to age, neither for infant ($p = 0.82$) or for mothers ($p = 0.60$). No gender effect was found.

Switching Pauses

The samples comprised 838 switching pauses lasting on average 730 ms ± 543.6 (range: 50–2974). Multiple regression analysis showed that switching pauses were on average 174.0 ms longer in samples involving older infants ($p = 0.007$). No significant gender effect was found ($p = 0.16$).

Among the switching pauses, 60.5% concerned pairs of vocalizations that were either between an IM pair of vocalizations (30.1%) or between a MI pair of vocalization (30.4%), i.e., switching pauses which did not involve an overlapping vocalization. Both types of switching pause were analyzed separately. Multiple regression analysis confirmed the overall age effect for all switching pauses ($p = 0.013$) and showed that switching pauses inside an IM pair of vocalizations were on average 135.4 ms shorter than those inside a MI pair of vocalizations ($p = 0.014$). This difference was not impacted by infant age ($p = 0.68$).

Turn-Taking Sequences

In total, 489 TTSs were identified, lasting on average 5.68 s ± 4.39 (range: 0.3–28.2) and ranging from 2 to 18 turns (mode of three turns, i.e., three alternating vocalizations). Because duration and number of turns were highly correlated [$r = 0.81$; $t(487) = 30.6$, $p < 0.0001$], both measures were analyzed separately in order to appreciate whether they were likely to be explained by age.

Multiple regression conducted on the number of turns per TTS did not reveal any significant effect, neither of age ($p = 0.57$) nor of gender ($p = 0.56$). However, we did observe a significant difference between both age groups regarding the duration of

TTSs: TTSs of older infants lasted 1.22 s longer ($p = 0.005$; see **Figure 3**). No significant gender effect was found ($p = 0.53$).

TTSs were divided into two subgroups depending on which partner initiated the sequence. Sequence initiation was coded when a vocalization occurring after a pause lasting at least 3 s was followed by at least one vocalization by the partner within less than 3 s. Among the 489 TTSs collected, 44.8% were initiated by infants. Logistic regression showed that TTS initiation by infant was most likely to occur in older infants ($p = 0.003$; $OR = 1.8$, 95% CI: 1.22 2.77). No significant gender effect was found ($p = 0.30$).

Turn-taking sequences were also divided in two subgroups according to which partner terminated the sequence. A sequence termination was coded when a vocalization following at least one vocalization by the partner within less than 3 s was followed by a pause of at least 3 s. Among the 489 TTSs observed, 37 (7.6%) ended with a co-vocalization and were then excluded from analysis. Twenty four percent of the remaining 447 TTSs were ended by infants. Logistic regression showed that TTSs were more likely to be ended by infants in older infants ($p = 0.004$; $OR = 2.0$, 95% CI: 1.26 3.23). We also observed a significant gender effect: boys were less likely to end a TTS than girls ($p = 0.009$; $OR = 0.55$, 95% CI: 0.35 0.86).

Discussion

In the present study, durational features of vocal interactions between mothers and infants from two age groups (2–3 and 4–5 months) were collected using acoustic analysis software. The aim of the study was, first, to ascertain whether vocal exchanges involving young infants has a turn-taking format and, second, to investigate developmental change in vocal turn-taking.

The first finding was that three out of four infant vocalizations are either followed or preceded by a maternal vocalization. Infants in both age groups vocalize in turn as frequently, and the most common turn-taking format at both ages is one involving three turns (two alternations between mother and infant). Indeed, infants respond to mothers' vocalizations as often as mothers respond to infants' vocalizations at both ages. Although the turn-taking format is observed in vocal interactions of infants aged between 2 and 5 months, the question that needs addressing is whether mothers are responsible for creating these formats or whether infants actively partake in shaping TTSs.

One way to investigate infants' active role in turn-taking is to ask whether they demonstrate anticipation of the mothers' turns and whether they show initiative in generating a vocal response from the mother. The present study provides partial answers to these questions because it is based on durational measures and does not include other acoustic features such as intonation contour, pitch matching or amplitude modulation. Some of our findings support the hypothesis that, already from the 2nd month of life, infants anticipate turns and initiate TTSs.

Infants at both ages used a high proportion of latched turns, that is they began to vocalize within a few milliseconds (less than 50 ms) of the end of the mother's vocalization, without any overlap. To our knowledge, no study of early MI vocal interaction has reported on the frequency or significance of latched turns. This finding is significant in teasing apart the roles of adults and infants in turn-taking exchanges. Indeed, although latched turns are less frequent than overlapping vocalization, it seems implausible that they would occur randomly at the frequencies we report. The temporal window for initiating vocalization is markedly smaller for latched turns than for turns involving a pause and for overlapping vocalizations. Indeed, the average time frame within which an infant vocalization is described

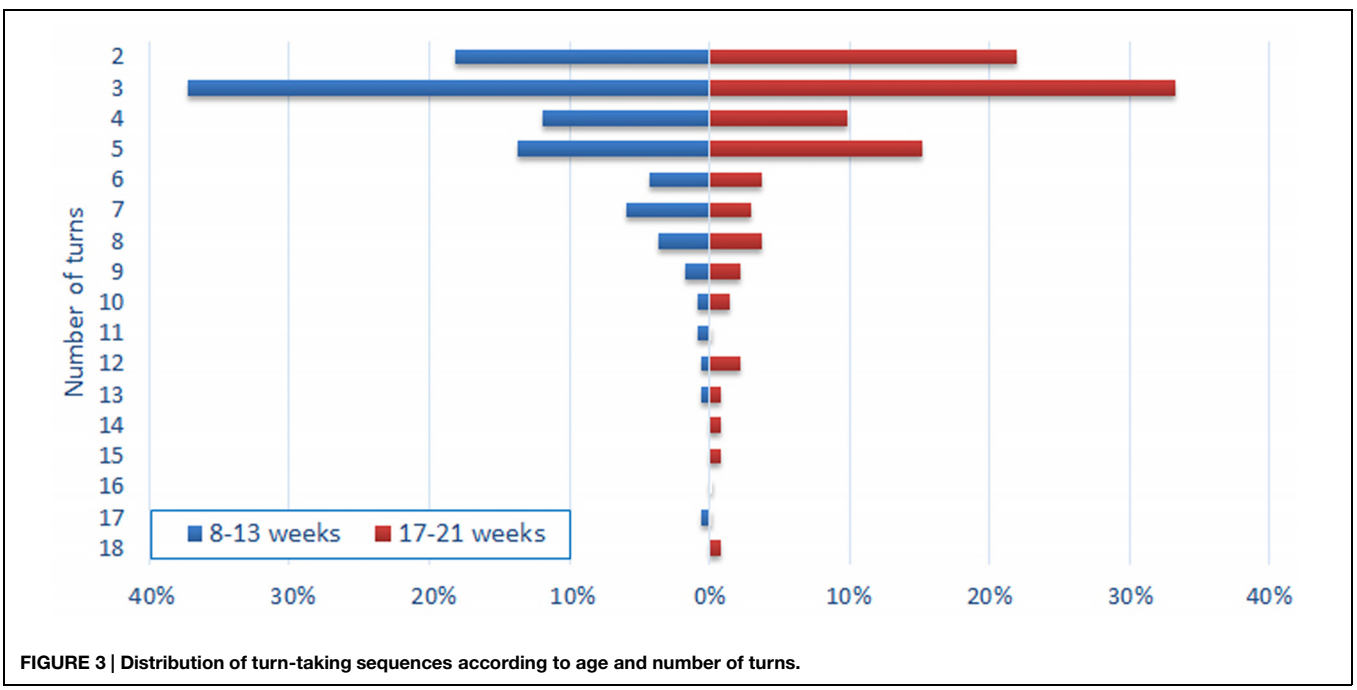


FIGURE 3 | Distribution of turn-taking sequences according to age and number of turns.

as an 'overlapping vocalization' in our study is 1145 ms (the average duration of a maternal vocalization) and the average time frame within which an infant vocalization is involved in a turn (excluding latched turns) is 730 ms (the average duration of a switching pause). The high frequency of latched infant vocalizations suggests that infants actively shape the turn-taking organization of an exchange and anticipate events within it. Based on a comparison of the probabilities of infants producing different prosodic contours, a previous study showed that 3-months-old selectively match their vocalizations to preceding maternal vocalization (Gratier and Devouche, 2011). Thus the low probability that an infant should vocalize at a particular moment in time or in a particular way can be interpreted as a high probability that the infant vocalizes with purpose.

The finding that infants in the 3rd months already frequently perform latched turns poses a major challenge. How do young infants project the end of a turn as precisely as adults do in verbal conversation, with 'no-gap-no-overlap' (Sacks et al., 1974), without any lexico-syntactic information? Much research has shown that the projection of turn completion relies principally on linguistic cues (De Ruiter et al., 2006). It is possible that infants rely on durational cues to project the end of the mother's turn and that mothers too rely on such cues to project the end of the infant's vocalization. Maternal speech to infants is known to be rhythmic, and maternal utterances are usually short and bound by pauses. In our study, as in previous studies (Jaffe et al., 2001), maternal vocalizations lasted around 1 s. We wondered whether infants might learn their mothers' specific time-signature which would enable them to predict with high precision when most of her utterances end. However, the duration of the vocalization to which an infant vocalization is latched does not appear to sufficiently explain the occurrence of latching. Maternal vocalizations that preceded a latched turn were only marginally longer that those preceding a switching pause. Another plausible explanation for the high percentage of latched turns is that mothers interrupt their utterances as soon as they perceive the infant is about to vocalize. Mothers may use cues such as in-breath or change in muscle tone to predict the onset of an infant vocalization. However, if it were the case that mothers interrupt their utterances rather than that infants predict the end of the mother's utterance, maternal utterances preceding a latched infant vocalization should have been found to be shorter than maternal utterances preceding a switching pause. It is worth investigating whether other acoustic features might afford turn latching, such as drop in fundamental frequency and/or intensity. A more complete analysis of acoustic features of vocalization might provide greater insight into this phenomenon.

Furthermore, it is interesting to explore the possible functions of latched turns in protoconversational exchange. The experience of a seamless transition between self-expression and other-expression may be highly relevant for a young infant, reinforcing an emerging sense of agency in the 1st months of life. Latched turns may be conceived as joint action, where each individual's actions are coordinated so as to achieve a joint outcome and where each individual's action cannot be understood in isolation from the others' (Sebanz et al., 2006).

Overall, comparison of the two age groups revealed both continuity and change in turn-taking organization. We do not find any differences in terms of mean frequency of vocalizations involving overlap between the two ages. Nor do we find that the duration of switching pauses decreases with age. On the contrary in our data they are longer among the older infants. Older infants do not perform proportionately more latched turns than younger infants. We did find that switching pauses were more frequent in younger infants' interactions. This finding should entail that younger infants partake in a greater number of alternating vocalizations but we did not find that the number of turns within TTSs increases with age. This inconsistency may be explained by the quantification method we used for turn-taking. Overlapping vocalizations and latched vocalizations were taken into account in the quantification of number of turns whereas the frequency of switching pauses was quantified on the basis of non-overlapping vocalizations alone. It is therefore not possible to describe the contribution of each type of vocalization within the TTSs. Future studies should be aimed at more precise description of overlapping vocalization in MI interaction.

Turn-taking sequences were found to be longer for older infants despite the fact that infant vocalizations have similar durations in both age groups and that mothers' vocalizations are shorter in interactions with older infants. Longer TTSs at 4–5 months than at 2–3 months are most likely explained by the increase in switching pause duration between the two ages. Furthermore, a greater number of infant vocalizations were either responded to or were responses to the mother among the younger infants, and older infants vocalized more often twice in succession. These two age-related changes may reflect a transition between the highly social cooing stage and the more solitary exploratory sound play that follows in vocal development (Oller, 2000). Hilbrink et al. (submitted) related their finding that turn-timing slows down around 9 months to the emergence of socio-cognitive skills that are crucial for communication and language, such as joint attention and cooperative cognition. In the present study too the younger infants responded to maternal vocalization faster than the older infants did. Similarly, this finding may be explained by a reorganization of social attention that is known to occur around 4 months, when infants become increasingly interested in object play and begin to show shared attention.

Our findings confirm the active role of mothers in adjusting their vocal behavior to infants. Mothers' vocalizations were found to be longer with younger infants, and they produce more successive vocal utterances when interacting with younger infants. Mothers perform more latched turns with older infants and switching pauses between IM pairs of vocalizations are shorter on average than between MI pairs, suggesting that mothers respond to infants faster than infants respond to mothers. In a previous study, the duration of switching pauses in vocal turn-taking between 2 and 3 months-old infants and mothers was found to vary cross-culturally in relation with parenting styles and cultural representations (Gratier, 2003). Switching pause duration may thus reflect mothers' representation of the infant's communicative ability more than the infants' turn-taking competence. We also found that mothers' vocalizations are longer with boys than with girls confirming

that gender-based representations impact infant-directed speech (Kitamura and Burnham, 2003).

The significance of our findings on initiation and termination of TTSs is ambiguous. We found that older infants were more likely to initiate a TTS. However, initiating a sequence does not imply that older infants show greater initiative in turn-taking, as this finding could be explained by mothers responding more often to the vocalizations of older than of younger infants. In fact, a recent study has shown that, without knowledge of infant age, naïve listeners implicitly attribute greater communicative intentionality to vocalizations of 5 months-old infants than to those of 1 month-old (Gratier et al., submitted). Our data also show that older infants more frequently terminate TTSs than younger infants. This may also be seen to reflect the mother's stance toward the infant rather than a change in the way infants organize turn-taking because mothers may not respond as often to terminal vocalizations with older infants. It would be interesting to take into account the acoustic characteristics of these initial and terminal infant vocalizations in order to gain insight into the question of infant initiative. Prosodic cues such as intonation contour may signal initiative to mothers, for both initiating and terminating turn-taking. It is also possible that mothers do not respond to terminal vocalizations because the TTS has reached a durational threshold, her unresponsiveness would then serve a regulatory function. It is worth investigating the transformations of the two partners' roles in turn-taking across ages. Indeed, mothers may progressively give the infant more and more prominence in the interaction, considering her more and more as an active partner with "something to say" while making the rules of turn-taking more salient. Infants may become less interested in a vocal turn-taking and more interested in multimodal turn-taking around object play or shared activities.

Overall our findings do not clearly support the hypothesis that infants' become more competent at turn-taking between the 3rd and 5th months of life. Indeed, we find neither shorter switching pause durations nor less overlap between these ages. Rather, our findings suggest that turn-taking organization is sensitive to changes in infants' social motives, reflecting a growing involvement in object play and joint activity. Although mothers are clearly highly adaptive and active in vocal exchanges with infants, we have found evidence that already very young infants play an active role in shaping the unfolding of TTSs. This exploratory study on the early development of turn-taking points to the possibility that turn-taking in preverbal interaction adapts to infants' changing motives for communicating and learning, and paves the road to the crucial socio-cognitive skills that precede and enable language use.

Acknowledgments

We would like to thank all the families who participated in this study. We especially thank Anne Bobin-Bègue and Giulietta Von Salis for help with data collection, doctors Bourrelier and Guillemot for providing access to MI health care services and Rose-Marie Guérinel for help contacting families. We also gratefully acknowledge the help of Ofer Tchernichovski on the use of Sound Analysis Pro. This research was supported by the Fondation de France (reference # 00015165) and by a CAPES grant (BEX-10506/12-8) from Brazil.

References

Bateson, M. C. (1975). Mother-infant exchanges: the epigenesis of conversational interaction. *Ann. N. Y. Acad. Sci.* 263, 101–113. doi: 10.1111/j.1749-6632.1975.tb41575.x

Beaumont, S. L., and Bloom, K. (1993). Adults' attributions of intentionality to vocalizing infants. *First Lang.* 13, 235–247. doi: 10.1177/014272379301303805

Beebe, B., Jaffe, J., Feldstein, S., Mays, K., and Alson, D. (1985). "Interpersonal timing: the application of an adult dialogue model to mother-infant vocal and kinesic interactions," in *Social Perception in Infants*, eds T. Field and N. Fox (Norwood, NJ: Ablex Publishing Corporation), 217–247.

Bloom, K. (1988). Quality of adult vocalizations affects the quality of infant vocalizations. *J. Child Lang.* 15, 469–480. doi: 10.1017/S0305000900012502

Bloom, K., and Lo, E. (1990). Adult perceptions of vocalizing infants. *Infant Behav. Dev.* 13, 209–219. doi: 10.1016/0163-6383(90)90031-3

Bloom, K., Russell, A., and Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *J. Child Lang.* 14, 211–227. doi: 10.1017/S0305000900012897

Bloom, L. (1990). "Developments in expression: affect and speech," in *Psychological and Biological Approaches to Emotion*, eds N. L. Stein, B. Leventhal, and T. Trabasso (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 215–246.

Caskey, M., Stephens, B., Tucker, R., and Vohr, B. (2011). Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics* 128, 910–916. doi: 10.1542/peds.2011-0609

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd Edn. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Cohn, J. E., and Tronick, E. (1987). Mother-infant face-to-face interaction: the sequence of dyadic states at 3, 6, and 9 months. *Dev. Psychol.* 23, 68–77. doi: 10.1037/0012-1649.23.1.68

Cossette, L., Malcuit, G., Pomerleau, A., and Julien, D. (1986). Temporal structure of maternal language directed at infants 3 months old. *Can. J. Psychol.* 40, 414–422. doi: 10.1037/h0080106

Delack, L. B., and Fowlow, P. J. (1978). "The ontogenesis of different vocalizations: development of prosodic contrastivity during the first year of life," in *The Development of Communication*, eds N. Waterson and C. Snow (London: Wiley), 93–110.

De Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

D'Odorico, L., and Franco, F. (1991). Selective production of vocalizations in different communication contexts. *J. Child Lang.* 18, 475–499. doi: 10.1017/S0305000900011211

Ginsburg, G. P., and Kilbourne, B. K. (1988). Emergence of vocal alternation in mother-infant interchanges. *J. Child Lang.* 15, 221–235. doi: 10.1017/S0305000900012344

Goodwin, C. (1986). Gestures as a resource for the organization of mutual orientation. *Semiotica* 62, 29–49. doi: 10.1515/semi.1986.62.1-2.29

Gratier, M. (2003). Expressive timing and interactional synchrony between mothers and infants: cultural similarities, cultural differences, and the immigration experience. *Cogn. Dev.* 18, 533–554. doi: 10.1016/j.cogdev.2003.09.009

Gratier, M., and Devouche, E. (2011). Imitation and repetition of prosodic contour in vocal interaction at 3 months. *Dev. Psychol.* 47, 67–76. doi: 10.1037/a0020722

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., et al. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monogr. Soc. Res. Child Dev.* 66, 1–8. doi: 10.1111/1540-5834.00137

Keller, H., Lohaus, A., Völker, S., Cappenberg, M., and Chasiotis, A. (1999). Temporal contingency as an independent component of parenting behavior. *Child Dev.* 70, 474–485. doi: 10.1111/1467-8624.00034

Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. *Infancy* 4, 85–110. doi: 10.1207/S15327078IN0401_5

Lavelli, M., and Fogel, A. (2013). Interdyad differences in early mother–infant face-to-face communication: real-time dynamics and developmental pathways. *Dev. Psychol.* 49, 2257–2271. doi: 10.1037/a0032268

Lipkind, D., Marcus, G. F., Bemis, D., Sasahara, K., Jacoby, N., Takahashi, M., et al. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* 498, 104–108. doi: 10.1038/nature12173

Magyari, L., and De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

Murray, L., and Trevarthen, C. (1985). "Emotional regulations of interactions between two-month-olds and their mothers," in *Social Perception in Infants*, eds T. M. Field and N. A. Fox (Norwood, NJ: Ablex Publishing Corporation), 177–197.

Murray, L., and Trevarthen, C. (1986). The infant's role in mother–infant communications. *J. Child Lang.* 13, 15–29. doi: 10.1017/S0305000900000271

Oller, D. K. (1980). "The emergence of the sounds of speech in infancy," in *Child Phonology*, Vol 1, eds G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (New York, NY: Academic Press), 93–112.

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Hove: Psychology Press.

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., and Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6318–6323. doi: 10.1073/pnas.1300337110

Papoušek, M., and Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Lang.* 9, 137–157. doi: 10.1177/014272378900900603

Rutter, D. R., and Durkin, K. (1987). Turn-taking in mother-infant interaction: an examination of vocalizations and gaze. *Dev. Psychol.* 23, 54–61. doi: 10.1037/0012-1649.23.1.54

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Smith, N. A., and Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy* 13, 410–420. doi: 10.1080/15250000802188719

Stark, R. E. (1980). "Stages of speech development in the first year of life," in *Child phonology*, eds G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (New York, NY: Academic Press), 73–90.

Stern, D. N., Beebe, B., Jaffe, J., and Bennett, S. L. (1977). "The infant's stimulus world during social interaction: a study of caregiver behaviors with particular reference to repetition timing," in *Studies in Mother–Infant Interaction*, ed. H. R. Schaffer (New York, NY: Academic Press), 177–202.

Stern, D. N., and Gibbon, J. (1979). "Temporal expectancies of social behaviours in mother-infant play," in *Origins of the Infant's Social Responsiveness*, ed. E. B. Thoman (Hillsdale, NJ: Lawrence Erlbaum Associates), 409–429.

Stern, D. N., Jaffe, J., Beebe, B., and Bennett, S. L. (1975). Vocalizing in unison and in alternation: two modes of communication within the mother-infant dyad. *Ann. N. Y. Acad. Sci.* 263, 89–100. doi: 10.1111/j.1749-6632.1975.tb41574.x

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Striano, T., Henning, A., and Stahl, D. (2005). Sensitivity to social contingencies between 1 and 3 months of age. *Dev. Sci.* 8, 509–518. doi: 10.1111/j.1467-7687.2005.00442.x

Tchernichovski, O. (2012). *User Manuel*. Availble at: http://soundanalysispro.com

Tchernichovski, O., Nottebohm, F., Ho, C. E., Bijan, P., and Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Anim. Behav.* 59, 1167–1176. doi: 10.1006/anbe.1999.1416

Trevarthen, C. (1977). "Descriptive analyses of infant communicative behavior," in *Studies in Mother-Infant Interaction*, ed. H. R. Schaffer (London: Academic Press), 227–270.

Trevarthen, C. (1979). "Communication and cooperation in early infancy: a description of primary intersubjectivity," in *Before Speech: The Beginning of Interpersonal Communication*, ed. M. Bullowa (Cambridge: Cambridge University Press), 321–347.

Trevarthen, C. (1993). "The self-born in intersubjectivity: the psychology of an infant communicating," in *The Perceived Self: Ecological and Interpersonal Sources of Self-Knowledge*, ed. U. Neisser (New York, NY: Cambridge University Press), 121–173.

Trevarthen, C., and Hubley, P. (1978). "Secondary intersubjectivity: confidence, confiding and acts of meaning in the first year," in *Action, Gesture and Symbol: The Emergence of the Language*, ed. A. Lock (London: Academic Press), 183–229.

Tronick, E., Als, H., Adamson, L., Wise, S., and Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *J. Am. Acad. Child Psychiatry* 17, 1–13. doi: 10.1016/S0002-7138(09)62273-1

Van Egeren, L. A., Barratt, M. S., and Roach, M. A. (2001). Mother–infant responsiveness: timing, mutual regulation, and interactional context. *Dev. Psychol.* 37, 684–697. doi: 10.1037/0012-1649.37.5.684

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 29 September 2015
doi: 10.3389/fpsyg.2015.01492

Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction

Elma E. Hilbrink[1], Merideth Gattis[2] and Stephen C. Levinson[1]*

[1] Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [2] School of Psychology, Cardiff University, Cardiff, UK

To accomplish a smooth transition in conversation from one speaker to the next, a tight coordination of interaction between speakers is required. Recent studies of adult conversation suggest that this close timing of interaction may well be a universal feature of conversation. In the present paper, we set out to assess the development of this close timing of turns in infancy in vocal exchanges between mothers and infants. Previous research has demonstrated an early sensitivity to timing in interactions (e.g., Murray and Trevarthen, 1985). In contrast, less is known about infants' abilities to produce turns in a timely manner and existing findings are rather patchy. We conducted a longitudinal study of 12 mother–infant dyads in free-play interactions at the ages of 3, 4, 5, 9, 12, and 18 months. Based on existing work and the predictions made by the Interaction Engine Hypothesis (Levinson, 2006), we expected that infants would begin to develop the temporal properties of turn-taking early in infancy but that their timing of turns would slow down at 12 months, which is around the time when infants start to produce their first words. Findings were consistent with our predictions: infants were relatively fast at timing their turn early in infancy but slowed down toward the end of the first year. Furthermore, the changes observed in infants' turn-timing skills were not caused by changes in maternal timing, which remained stable across the 3–18 months period. However, the slowing down of turn-timing started somewhat earlier than predicted: at 9 months.

Keywords: turn-taking, mother–infant interaction, infants, timing, communicative development

OPEN ACCESS

Edited by:
Marcela Pena, Pontifical Catholic University of Chile, Chile

Reviewed by:
Marilyn Vihman, University of York, UK
Veronica Figueroa, Goethe-Universität Frankfurt am Main, Germany

****Correspondence:***
Elma E. Hilbrink, Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525 XD, Netherlands
elma.hilbrink@mpi.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *21 November 2014*
Accepted: *16 September 2015*
Published: *29 September 2015*

Citation:
Hilbrink EE, Gattis M and Levinson SC (2015) Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. Front. Psychol. 6:1492. doi: 10.3389/fpsyg.2015.01492

Introduction

For a conversation to run smoothly, a tight coordination of interaction between speakers is required. In their seminal paper on the organization of turn-taking in conversation Sacks et al. (1974) noted that in conversation mostly one speaker talks at a time, that occurrences of overlap (i.e., more than one speaker at the time) are common, but brief, and that the vast majority of turn transitions (i.e., the switch from one speaker to the next) are characterized by either no gap and no overlap or by a slight gap or slight overlap. Moreover, a recent study comparing across a variety of languages demonstrated that this close-timing of turns might be universal (Stivers et al., 2009). Stivers et al. (2009) compared the turn transitions in naturalistic conversation across 10 diverse languages that differed, amongst other things, in word order, sound structure, and grammatical options. They found that, despite of some variation in the overall distributions, turn transitions in all of the languages had a mode between 0 and 200 ms. This is remarkably fast considering that it

takes at least 600 ms to plan a short turn at talk (Levinson, 2013). This suggests that in order to produce smooth transitions from one speaker to the next, speakers need to simultaneously listen to the other speaker, plan their own turn, and predict when to launch that turn. It therefore comes as no surprise that children who are still in the process of acquiring language are much slower than adults. Garvey and Berninger (1981), for example, reported a mean gap duration of 1100–1800 ms in preschoolers engaged in child–child conversation. Casillas et al. (under review) reported a median gap duration in mother child question–answer pairs of 587 ms for children between the ages of 2;8 and 3;5.

The Interaction Engine hypothesis suggests that the infrastructure for this remarkably tight coordination underlying conversation is based on our sensitivity to the timing of turns and our ability to anticipate and recognize others' communicative intentions (Levinson, 2006). Moreover, this social interactional infrastructure is thought to be present early in infancy, before infants have acquired language, and is hypothesized to be the foundation for communicative turn-taking. Thus, even though preschool-aged children have been shown to be slow compared to adults, the Interaction Engine hypothesis suggests that preverbal infants already have (parts of) the necessary interactional infrastructure at their disposal. More specifically, the Interaction Engine hypothesis suggests that the temporal aspect of this infrastructure, i.e., turn-timing, develops early in infancy. Like scholars such as, e.g., Bruner (1975, 1981) and Bateson (1975) this hypothesis views early mother–infant exchanges as proto-conversations and places great emphasis on the role of early interactional skill. But, while Bruner focuses mostly on speech acts, or understanding of communicative intentions, the Interaction Engine hypothesis ascribes important roles to both the understanding of communicative intentions as well as the temporal aspect of these early exchanges. The present paper aims to explore the development of the temporal aspect of turn-taking, i.e., turn-timing, during infancy in naturalistic interactions.

Infants spend a substantial portion of their awake-time in face-to-face interactions with their caregivers and it is this face-to-face conversational setting that provides an important context for infants to start acquiring language. Moreover, well-before infants have acquired language they start to interact in social exchanges characterized by turn-taking patterns, such as peek-a-boo games and give and take sequences (e.g., Bates et al., 1975; Ratner and Bruner, 1978; Rochat et al., 1999). Thus, it is in these face-to-face interactions where one might expect to observe infants' earliest communicative abilities. Indeed, Kaye suggested for example that the burst-pause patterns observed in sucking during feeding and in facial expressions in mother–infant face-to-face interactions, resemble the turn-taking patterns in communication and could be the basis for acquiring communicative turn-taking abilities (Kaye, 1977; Kaye and Fogel, 1980). Furthermore, a recent study on face-to-face interaction in the first 6 months of life showed that mother and infant spend most of their time in unilateral communication, i.e., where mother tries to engage the infant but the infant is not attending, however, with increasing infant age the time they spend in symmetrical communication, i.e., being mutually engaged with a joint focus of attention, increases (Hsu and Fogel, 2003). These mutually engaged interactions can consist of behaviors in multiple modalities such as vocalizations, gaze and smiles, which have been shown to be temporally ordered. For example, Hsu et al. (2001) found that infants' speech-like sounds occurred more often when their mothers were smiling, when infants were looking at their mothers, and also when they themselves were smiling. Moreover, speech-like vocalizations, compared to non-speech-like vocalizations, were more likely to be preceded by maternal smiling, indicating a temporal ordering of these interactional behaviors.

While the above studies demonstrate the existence of a general temporal coordination, they do not allow for the fine-grained temporal analyses that are common in adult studies on timing of turns in conversation (e.g., ten Bosch et al., 2005; Stivers et al., 2009). In order to assess this precise timing of turns in mother–infant interactions in the present study we specifically focused on vocal exchanges. Various studies on vocal exchanges in naturalistic interactions suggest that these early vocal exchanges between mothers and infants can be described as conversation-like (Bateson, 1975; Snow, 1977). Furthermore, several experimental studies, in which it was manipulated whether infants received contingent responses (i.e., responses related to the infants' behavior) or yoked (non-contingent) responses, have demonstrated that contingent responding has positive effects on the quality of infant vocalizations and facilitates turn-taking behavior in vocal exchanges between mothers and infants (e.g., Bloom et al., 1987; Bloom, 1988; Masataka, 1993; Goldstein et al., 2003). However, these studies are solely based on experimental manipulation and only assess short-term effects. Nevertheless, various scholars have suggested an important role for early turn-taking behaviors in the development of language. For example Ginsburg and Kilbourne (1988) state: "*The possibility has become widely accepted over the past decade that the temporal patterning of social interchanges involving the young infant has important implications for linguistic development*" (p. 221). Thus infants' (vocal) turn-taking behaviors in face-to-face interactions are suggested to be among the earliest communicative abilities that infants might demonstrate.

Some experimental evidence indicates that infants are sensitive to the timing of turn-taking in social exchanges. Striano et al. (2006) adapted a setup originally designed by Murray and Trevarthen (1985), in which mother and infant interact via screens, by adding a device to the video setup that allowed the ongoing interaction to be delayed by 1 s. In the original setup infants saw a live interaction and a replay of the interaction, while in the adapted version the interaction was delayed (online) by only 1 s. Similar to Murray and Trevarthen (1985), Striano et al. (2006) found that 3- and 6- month-olds gazed less at their mothers during the delayed interaction compared to the live interaction. In addition, an eye tracking study by Thorgrimsson (2014) has shown that when 1-year-olds observe two people in a face-to-face context, they are quicker to shift their gaze to person (B) when person (A) uttered a sentence compared to when person (A) emitted a non-speech sound (e.g., throat clearing, singing). In both conditions person (B) never responded, thus infants could

not have learned what to expect during the experiment. This indicates that 1-year-olds expect speech to provoke a response. Several eye-tracking studies by Casillas and Frank (2013) have taken this paradigm one step further and have demonstrated that 1-year-olds are able to anticipate the upcoming turn when observing two adults or two puppets having a conversation. Together, these studies demonstrate that from early on in the first year, before infants have acquired language, infants are sensitive to changes in the timing of social interactions.

In contrast to infants' perception and comprehension of timing, less is known about infants' abilities to produce turns in a timely manner, and existing findings are rather patchy. A few studies have tried to assess turn-timing structure in infancy. Bateson (1975), for example, provided a detailed description of five interactions between one mother and her baby recorded between the ages of 1.5 and 3.5 months. Her analyses suggested that mother and infant alternated turns and that this alternation of turns seemed mutual. Both mother and infant left longer silences between two consecutive utterances made by themselves compared to when they responded to each other's utterances; however, this pattern only reached significance for the mother. Naturalistic studies further suggest that infants start out producing a large portion of their vocalizations in overlap with their mothers' utterances. The amount of overlapping vocalizations is said to decrease in favor of a more alternation-like pattern from around 4 months of age (Elias et al., 1986; Ginsburg and Kilbourne, 1988; Elias and Broerse, 1996). However, a study by Rutter and Durkin (1987) assessing turn-timing from 9 to 36 months reported an increase in overlapping vocalizations from 9 to 24 months. In addition, studies assessing precise timing have also resulted in mixed findings. **Table 1** summarizes several of these earlier studies and their methods. The mean gap durations reported between the ages of 1 and 4 months range from 800 to 1370 ms (Bateson, 1975; Elias et al., 1986; Beebe et al., 1988). Jasnow and Feldstein (1986) reported a gap duration for 9-month-olds of 875 ms. Whether one would conclude from this that infants remain stable in their timing or start to speed up at 9 months depends on which of the studies you rely on for the gap durations at 1–4 months. To complicate things further, studies have differed on how to record the timing. Bateson (1975) for example reported maternal and infant gap durations which were recorded from the onset of the other's utterance to the onset of their own utterance, i.e., these included not just the transition from one speaker to the next but also the duration of the previous utterance. Others chose to look at the actual silence between two utterances. In other words they measured the time between the end of the utterance of one speaker to the beginning of the utterance of the next speaker (Elias et al., 1986; Jasnow and Feldstein, 1986; Beebe et al., 1988). Together, these findings demonstrate that the developmental picture of infant turn-timing is far from clear. Furthermore, previous findings are mostly based on only one age group or on cross-sectional samples. Longitudinal designs tracking development over extended periods of time could provide valuable insights about stability and developmental change in turn-timing skills. A few longitudinal studies exist, but these studies are based on small samples including 1–3 children and/or cover a short period of time, i.e., 3–5 months (e.g., Bateson, 1975; Snow, 1977; Ginsburg and Kilbourne, 1988).

TABLE 1 | Summary of several studies assessing precise turn-timing in infancy.

Study	Bateson, 1975	Elias et al., 1986	Jasnow and Feldstein, 1986	Beebe et al., 1988
Age (in months)	1.5–3.5	3–4	9	4
N	1	6	29	15
Average gap duration in ms (infant)	1370*	1200	875	800
Average gap duration in ms (mother)	1430*	750	775	700

**Includes preceding utterance duration.*

In addition to the questions about the developmental trajectory, considerable debate exists with respect to whether these vocal exchanges between mothers and their infants are reciprocally structured. Or whether in fact infants are randomly vocalizing while mothers are responsible for establishing a turn-taking structure, with possible observed changes due to changes in maternal behavior. Snow (1977) suggested that mothers are mainly responsible for maintaining the conversational structure. Anderson et al. (1977) found evidence for reciprocity in vocal exchanges at 3 months of age, while Rosenthal (1982) observed reciprocity in vocal interactions of neonates and their mothers. Furthermore, Jasnow and Feldstein (1986), Beebe et al. (1988), Jaffe et al. (2001) conducted a series of studies to assess mothers' and infants' capacity for interpersonal accommodation to gap durations (i.e., whether the moment of silence between speaker transitions is sensitive to the partner's timing). The findings of these studies suggest that by 9 months of age infants' gap durations were influenced by their mothers' gap durations and vice versa. Jasnow and Feldstein (1986) called this interpersonal accommodation. Furthermore, they found that 4-month-olds change the length of their gap durations depending on who they interact with. For example, 4-month-olds left longer pauses when interacting with their mothers compared to when they were interacting with a stranger. This finding suggests that, contrary to Snow (1977), even 4-month-olds might be accommodating their gap durations.

There remains then considerable uncertainty about the development of turn-timing in preverbal infants. Especially with regards to their abilities to produce turns in a timely fashion and to reciprocally structure vocal exchanges. Existing findings on mother–infant vocal turn-timing are fragmented and mostly cross-sectional. Research, and especially longitudinal research, that tracks the development of turn-timing from early in infancy until the ages at which infants first start to produce language is still lacking. Moreover, studies have focused on assessing either overlapping vocalizations or precise gap durations but not both. The existing findings on the amount of overlapping vocalizations suggests a possible early decrease of overlapping vocalizations, while the various reports on infants' gap duration suggests either stability of gap durations across age or a decrease in gap

durations. This highlights the possibility that the developmental pattern for the amount of overlap might be different from the developmental pattern for gap durations. Therefore, to obtain a complete developmental picture research is needed that looks at both overlap and gaps. In studies of adult turn-timing overlap and gap measurements are often combined in a single floor-transfer offset measure where overlaps are treated as negative gaps, on the presumption that adult speakers are aiming at close transition times and may inadvertently come in early. However, in studies of infant turn-timing we cannot make the same assumption. Thus, in addition to the possibility that overlap and gap might show different developmental patterns, we can also not assume that infants, like adults, aim at close-transition times. Therefore, the present study set out to explore the development of turn-timing in a longitudinal study of mother–infant interaction between the ages of 3- and 18- months, by assessing, in contrast to earlier developmental studies, both gaps and overlaps. But, contrary to studies on adult turn-timing, we analyzed the overlap and gap durations as separate measurements.

Specifically, the present study aimed to describe the developmental pattern of infants' productive turn-timing abilities, including overlaps, gaps and within-turn pauses, i.e., the silence between two consecutive parts of the same turn of one interlocutor (see **Figure 1** for definitions). Furthermore we aimed to assess whether the observed mother–infant turn-timing patterns were reciprocally structured. We therefore conducted a longitudinal study of 12 mother–infant dyads in free-play interactions at six ages between 3 and 18 months. Based on previous work we expected that infants would begin to develop the temporal properties of turn-taking early in infancy. However, based on earlier work with older children and due to the complex nature of achieving smooth turn transitions we expected that at 12 months, i.e., around the time of the onset of language production, infant turn-timing would slow down (Garvey and Berninger, 1981; Casillas et al., under review). These predictions are consistent with the predictions of the Interaction Engine Hypothesis. This hypothesis suggests that preverbal infants acquire the temporal properties of conversational turn-taking early in infancy and that once infants start using language their turn-timing will slow down due to the need to integrate their developing linguistic skills with the existing interactional timing skills (Levinson, 2006). Furthermore, based on earlier work by Anderson et al. (1977), Jasnow and Feldstein (1986), Beebe et al. (1988), Jaffe et al. (2001) we expected that the observed turn-timing patterns were reciprocally structured at all ages.

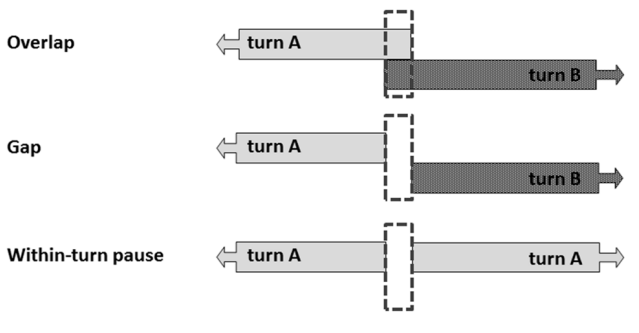


FIGURE 1 | A turn transition can consist of an overlap, a gap or if no turn transition occurs, a within-turn pause can occur.

Materials and Methods

Participants

For the present study free-play recordings from 12 mother–infant dyads (seven female) were analyzed when the infants were 3, 4, 5, 9, 12, and 18 months. Infants were originally recruited as part of the First Steps longitudinal study (Ellis-Davies et al., 2012). First Steps followed 39 healthy infants (18 female) from birth to 18 months. The study consisted of monthly testing on a variety of measures, free-play observations, diary data and questionnaires from 2 months onward. Infants were born to full gestation. All procedures for data collection were reviewed and approved by the Southeast Wales Research Ethics Committee of NHS Wales. In addition, all procedures and data analyses used in the present study were approved by the Radboud University Ethics committee under the research program: *INTERACT-Developmental studies* (Hilbrink and Levinson; project code ECG2012-2711-065). Written consent was obtained from all parents before the start of the study. Parents' level of education ranged from secondary school to postgraduate. Demographics on the full sample are available in Ellis-Davies et al. (2012). The level of education and maternal age of the 12 mothers in the present sample are comparable to the level of education and maternal age of the full sample (see **Table 2**). Parents were recruited during pregnancy through newspaper and web announcements and local events for expecting parents. The ages analyzed in the present study were chosen based on previous literature indicating an important transition around 4 months of age from vocalizing in overlap to a more turn-taking like pattern (Ginsburg and Kilbourne, 1988) and on studies suggesting important changes in communicative development emerging at 9 months (Bakeman and Adamson, 1984; Carpenter et al., 1998). In addition, ages were also chosen based on the

TABLE 2 | Demographic characteristics of the total sample and the current sample.

Maternal characteristics	Total sample (%)	Total sample *N*	Current sample (%)	Current sample *N*
Age in years at recruitment into the study				
16–20	0	0	0	0
21–25	11.11	4	8	1
26–30	25.00	9	25	3
31–35	33.33	12	33	4
36–40	30.56	11	33	4
Highest level of education attained				
High school	22.22	8	33	4
Undergraduate degree	52.78	19	50	6
Postgraduate degree	25.00	9	17	2

predictions made by the Interaction Engine Hypothesis that the temporal properties of turn-taking are realized early in infancy and are expected to change once infants start to acquire language (Levinson, 2006).

Procedure and Design

Mothers brought their infants to a 'mum and baby breakfast' at a community center or the university. For the present study only the recordings of the 10-min free-play mother–infant interactions were used. After mother and infant were seated mothers were asked to play with their infants as they would do at home. Although toys were available, mothers were not specifically instructed to use them. The experimenter would then leave the room and return after 10 min had passed.

Apparatus

The free-play interactions took place in a quiet room inside a tent, at 3, 4, and 5 months, or in a playpen, at 9, 12, and 18 months. Both setups were adapted to the needs of the infants at the various ages. To create privacy and consistency of visual environment, the mother–infant interactions at the younger ages (i.e., 3, 4, and 5 months) were conducted in a colorful tent. Infants were seated in a baby seat and the mother sat facing the infant. Three baskets containing age-appropriate toys, including soft toys, books and rattles were provided. The interactions were recorded using two static cameras, one filming the mother and the other filming the infant. Two microphones (AKG C1000S) recorded the sound from the same location as the two static cameras. A third camera was mounted on the mother's head with a headband allowing us to capture where the mother was looking. The signals of each of the three cameras were combined using a quad splitter, which resulted in a single time-synced split-screen video record.

At the older ages (i.e., 9, 12, and 18 months) the set-up was similar but because infants were able to sit upright and to move somewhat more, the tent was replaced by a playpen. This setup allowed infants to sit upright in a supportive seat within reach of the three baskets with toys. Interactions in the playpen were recorded using four static cameras: one capturing the infant, one capturing the mother, one capturing both mother and infant, and one capturing a bird's eye view of the playpen. The signals from each of these cameras were combined by a quad splitter, which resulted in a single time-synced split-screen video.

Transcription

The mother–infant interaction recordings of 12 mother–infant dyads were transcribed at six time points: 3, 4, 5, 9, 12, and 18 months of age using ELAN video annotation software (Sloetjes and Wittenburg, 2008). The 10-min recordings were transcribed for all maternal speech and for all infant sounds. Grunting, distress sounds and involuntary sounds, such as hiccups and sneezes, were excluded from analyses (e.g., Hsu et al., 2001). Maternal responses to involuntary infant sounds were, however, included as mothers often treated these sounds as communicative.

To calculate interrater reliability two recordings at each age were transcribed by another transcriber. With regards to the number of vocalizations identified for the infants the intraclass correlation (ICC) was 0.81. The percentage of agreement of a vocalization being a vocalization and not for example a distress sound was 76%. With regards to the timing of a vocalization, coders had to agree within a time window of two frames, i.e., 80 ms. The percentage of agreement for the time at which a vocalization started or ended was calculated based on all the vocalizations the coders agreed on being a vocalization. The agreement between coders for the time at which a vocalization started was 92% and the percentage of agreement for when a vocalization ended was 82%. For the number of utterances made by the mothers the ICC was 0.82. The percentage of agreement of an utterance being an utterance was 95%. The percentage of agreement for when an utterance started was 91% and for when an utterance ended was 86%.

Results

All turn transitions, both transitions from mother to infant and from infant to mother, were extracted from the transcriptions. A turn transition was defined as any switch from a maternal utterance to an infant vocalization or vice versa (see **Figure 1**). This resulted in 8555 turn transitions. As can be seen in **Table 3**, some individual variation exists with regards to the number of turn transitions across dyads at each age. An infant gap was defined as the gap between a maternal utterance and a vocalization from the infant, i.e., the onset of an infant vocalization minus the offset of the preceding maternal utterance. Maternal gap was measured in a similar way: the onset of the maternal utterance minus the offset of the preceding infant vocalization. Infant overlap was defined as transitions in which the infant started to vocalize when the mother had not yet finished speaking and maternal overlap was defined as whenever the mother started speaking when the infant had not yet finished vocalizing. Infant overlap was measured in the same way as infant gap durations but resulted, because of the overlap, in negative durations. Similarly, maternal overlap was measured in the same way as maternal gap duration. Furthermore, the moments

TABLE 3 | Number of turn transitions per dyad at each age.

	Age (in months)						Average across age
Dyad	**3**	**4**	**5**	**9**	**12**	**18**	
1	107	108	85	123	225	273	154
2	224	213	127	34	130	146	146
3	119	142	40	46	97	150	99
4	184	135	96	111	91	98	119
5	41	71	58	56	42	112	63
6	197	111	56	48	148	109	112
7	36	66	92	100	41	208	91
8	129	149	118	83	196	140	136
9	125	158	275	77	147	221	167
10	70	155	179	24	35	175	106
11	88	111	152	184	142	164	140
12	90	54	81	50	91	196	94
Average across infants	118	123	113	78	115	166	

of silence between two consecutive utterances or vocalizations by the same interlocutor, i.e., without the other interlocutor producing a turn in between, were also assessed. These within-turn pauses were measured in the same way as infant or maternal gap durations: the onset of an utterance or vocalization minus the offset of the preceding utterance or vocalization (see also **Figure 1** for the definitions).

To study the development of turn-timing in infancy we assessed three aspects of turn-timing. First we assessed the development of timing with regard to overlap. Next we assessed the development of infants' ability to time turns with regard to gap durations. We analyzed these separately (unlike many studies of adult turn-timing) because the developmental trajectory of overlapping turn transitions might differ from the developmental trajectory of transitions containing gaps. The third and final aspect of turn-timing we assessed was the development of turn-timing as whole, (i.e., all turn transitions, overlaps, and gaps) to assess whether infants' turn-timing differed from what would be expected if infants were vocalizing randomly in each age group.

In addition to studying the developmental trajectory of infants' turn-timing skills, we also explored whether mother–infant turn-timing patterns were reciprocally structured or whether changes in maternal behavior could account for possible changes in infant turn-timing with age. Analyses were conducted in R Development Core Team (2012) using the LME4 package (Bates et al., 2012). For the linear mixed effect modeling we followed the same procedures used by Hoicka and Akhtar (2011) and Hilbrink et al. (2013). All effects are expressed as odds ratios; when the odds ratio of an event is greater than one, the event is more likely to happen than not. The dependent variables that were included in the models were duration of overlap, duration of gap, or duration of within-turn pause. The variables that were included in the models as fixed effects were infant age in months (3, 4, 5, 9, 12, 18) and whether the overlap or gap durations were produced by mother or infant, i.e., person. Infant ID was included as random effect.

Overlap

Figure 2 shows the percentage of overlap at each age. At 3, 4, and 5 months infants produce just over a third of their vocalizations in overlap with their mothers. Mothers, produce between 14 and 21% of their turns in overlap with their infants at these same ages. However, by the time infants are 18 months infants have decreased the amount of overlap to similar levels as their mothers: to roughly 20% of their turns.

To assess whether infants also decreased the durations of their overlapping vocalizations we calculated the median durations of overlap (total number of overlapping turns: infant = 1180, mother = 840). Studies on adult turn-timing have demonstrated that brief overlaps occur often and that adults aim to launch their turn at the end of the previous turn which can result in brief overlap (Sacks et al., 1974; Jefferson, 2004). Thus if infants start decreasing the durations of overlapping turns this could indicate that they, like adults, are aiming for the end of the previous turn. **Figure 3** shows the infant and maternal median durations of overlap for each age. To assess whether infants and their mothers significantly change their duration of overlap we assessed all overlapping turn transitions using linear mixed modeling in R with duration of overlap as dependent variable and infant age and person (mother, infant) as fixed factors. We first build a base model with duration of overlap as dependent variable and infant ID as random effect. We compared this base model to models including age, person and an interaction of age and person. The base model was improved by adding age, $\chi^2(1) = 18.97$, $p = 0.000$, person, $\chi^2(1) = 51.55$, $p = 0.000$, and an interaction of age by person, $\chi^2(1) = 10.27$, $p = 0.000$. The final model included a significant interaction effect of age by person (model: $loglikehood = -16402$, $N = 2020$), $OR = 1369895094$, $p = 0.001$. To follow up this interaction we created separate models for mothers and infants. The separate analyses of data including only the overlap durations of the infants revealed no significant effect of age. The analyses on the data only including the maternal overlap durations did reveal a significant effect of age [$\chi^2(1) = 13.93$, $p = 0.000$; model: $loglikehood = -6865.8$]. Maternal overlap durations became significantly shorter with increasing infant age ($OR = 366679967$, $p = 0.000$).

Gap

To assess whether and how gap durations changed over time, infant and maternal median gap durations were calculated for each age (total number of gaps: infant = 2563, mother = 3992).

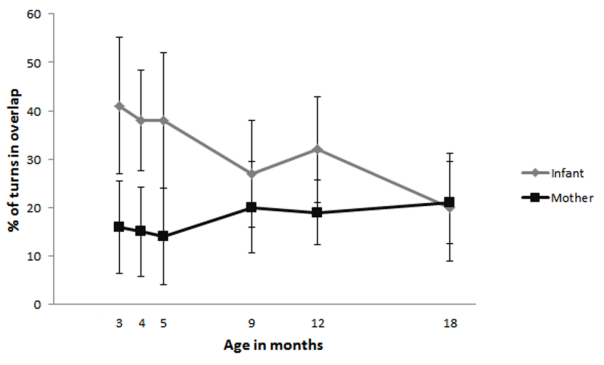


FIGURE 2 | Percentage of turns produced in overlap by infants (gray) and mothers (black).

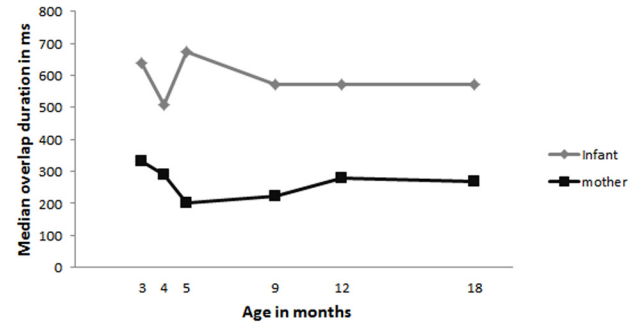


FIGURE 3 | Median durations of overlap for infant (gray) and mother (black). The closer the median is to zero the shorter the overlapping turn.

See **Figure 4** for the median durations at each age and **Table 4** for the range of median gap durations observed at each age. To assess whether these gap durations changed significantly across age we used linear mixed modeling with duration of the gaps as dependent variable and age and person (mother, infant) as fixed factors. We first build a base model with duration of gap as dependent variable and infant ID as random effect. We compared this base model to models including age, person and an interaction of age and person. The base model significantly improved by adding person, $\chi^2(1) = 27.34$, $p = 0.000$, an interaction of person by age, $\chi^2(2) = 1.23$, $p = 0.04$ and a trend was found for age, $\chi^2(1) = 3.26$, $p = 0.07$. The final model (*loglikehood* = −58435, $N = 6555$) contained a significant effect of age ($OR = 2987657$, $p = 0.01$), namely, gap durations became significantly larger with increasing age, and a significant age by person interaction, $OR = 1.246925e\text{-}08$, $p = 0.02$).

To follow up on the significant age by person interaction, we created separate models for mothers and infants. The separate analyses of data including only the gap durations for mother did not reveal a significant effect of age. The analyses on the data only including infant gap durations did reveal a significant effect of age [$\chi^2(1) = 5.95$, $p = 0.01$; model: *loglikehood* = −22928, $N = 2563$]. The effect of age indicates that infants' gap durations increase with age ($OR = 5025322$, $p = 0.01$). Furthermore, inspection of the individual data of the infants showed, compared to the earlier ages, that all infants slowed down at some point between 9 and 12 months, with most of the infants being the slowest at 9 months (8 out of 12 infants). Based on the effect of age, the findings in **Figure 4** and our observations in the individual data we conducted exploratory follow-up analyses to assess the difference in infant gap durations between 5, 9, 12, and 18 months. To do this we ran the same linear mixed model analyses on infant gap durations but separately on data sets including only the data at 5 and 9 months of age, 5 and 12 months of age and 5 and 18 months of age. These analyses revealed an effect of age for the dataset including 5 and 9 months [$\chi^2(1) = 25.18$, $p = 0.000$; model: *loglikehood* = −5716.9, $N = 630$] and the dataset including 5 and 12 months [$\chi^2(1) = 13.31$, $p = 0.000$; model: *loglikehood* = −7001.3, $N = 778$]. No effect of age was found for the dataset including the infant gap durations at 5 and 18 months. Thus infant gap durations are significantly longer at 9 months compared to 5 months ($OR = 2.806199e{+}96$, $p = 0.000$) and at 12 months compared to 5 months ($OR = 1.724077e{+}33$, $p = 0.000$), but not at 18 months compared to 5 months.

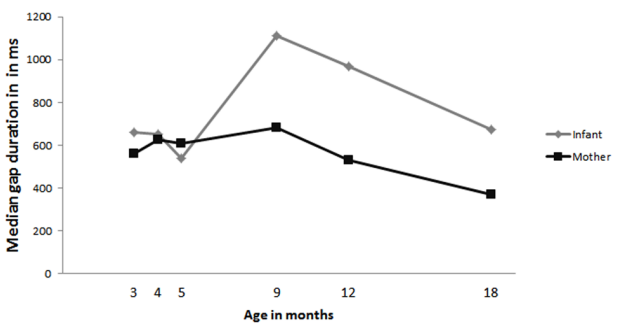


FIGURE 4 | Median gap durations for infant (gray) and mother (black).

Reciprocity: Do Infants Equally Structure the Interaction?

Thus far the results indicate that infants get better in producing their turns with less overlap (i.e., they decrease in the amount of overlap produced) as they get older, and that infant gaps are relatively short at 3, 4, and 5 months, but increase with age. However, it is possible that the changes observed in infant timing are not due to infants' changing turn-timing skills. Instead infants could be randomly producing turns, while mothers are trying to maintain a turn-taking structure by carefully timing their turns. For example mothers could change, as infants get older, in how long they wait for a vocalization from the infant. Especially when infants are older, mothers might expect a turn from their infant and might therefore increase their pauses between two consecutive utterances. Therefore we assessed, first, whether infants timed their turns significantly differently from what would be expected if they were vocalizing randomly at each age and, secondly, whether mothers changed their pause duration between two consecutive utterances.

To assess whether infants were randomly producing turns, we compared the observed distribution of infants' turn transitions (i.e., including both gaps and overlaps) to random distributions. The random distributions of possible infant turn transitions were estimated by looking at points in the interaction when the infant could have taken a turn, but did not. Transitions were identified where utterance (A) came from the mother and then the next utterance (B) also came from the mother, without the infant taking a turn, i.e., maternal within-turn pauses. An infant producing turns at random points could have taken a turn within a window of time from the start of the mother's turn to the end of the gap (any earlier or later would mean infants were transitioning from a different turn). Thus, the maternal turns were kept fixed, while the infants' turn onsets varied randomly. The distribution was built by calculating the height of the distribution at a given distance from the end of the mother's turn as the proportion of windows that included the given point. Because the time window, from the start of

TABLE 4 | Range of median gap durations at each age in ms.

	Month 3	Month 4	Month 5	Month 9	Month 12	Month 18
Infant	345.5–902.5	326–921	323–1408	542.5–3297	615–1872	485–1270
Mother	372.5–905.5	412–938	391–1204	445–1005	210–817	135–1145.5

the mother's turn until the end of the gap, always included the end of the maternal utterance, i.e., zero gap, the distribution has a peak around zero (see **Figure 5**). We then compared this 'random infant' distribution to the observed sample of infant turn transitions. This was done using a permutation test: i.e., the observations were randomly swapped between the observed sample and a random sample and then the difference in medians between these two new samples was calculated. A 1000 random infant samples were generated and permuted 10,000 times each with the observed sample. If the samples would have come from the same distribution, then the differences in medians should be normally distributed around zero. The probability of the given sample coming from the random distribution was calculated as the proportion of permutations resulting in a larger or equal difference in medians between the permuted samples than the actual difference, i.e., between the unpermuted samples. This probability was less than 1/10000000. In other words, out of 10 million permutations, none produced a difference in medians larger than the actual difference, suggesting that the actual distributions are significantly different ($p < 0.00000001$). Thus, at each age the infant turn transitions observed in our dataset differed significantly from the random distributions.

Next we analyzed whether mothers changed their within-turn pause durations, i.e., the silence between two consecutive utterances, to allow their infants more time to produce a turn. Maternal median within-turn pause durations did not change with increasing infant age (see **Figure 6** for median within-turn pauses and **Table 5** for the range in medians observed). However, as can be seen in **Figure 6**, mothers leave longer pauses after their own utterances (within-turn pause duration) compared to when they respond to an infant vocalization (switch), a pattern

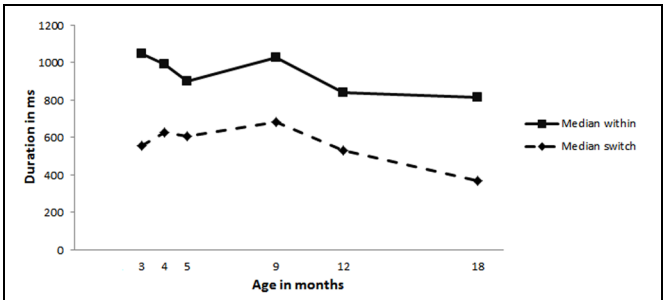


FIGURE 6 | Mothers' median within-turn pause durations (solid line) and median gap durations after an infant vocalization (dashed line).

found in adult–adult interaction. Analyses conducted on the infant within-turn pause durations revealed a significant effect for age [$\chi^2(1) = 27.65$, $p = 0.000$; model: $loglikehood = -15068$, $N = 1699$]. The within-turn pause durations of the infants significantly increase with age, $OR = 5.950459e{+}18$, $p = 0.000$, indicating they wait longer after their own turn as they get older. Furthermore, as can be seen in **Figure 7**, infants do not yet leave longer gaps after their own vocalizations compared to after their mothers' utterances.

General Discussion

As reported above, we conducted a longitudinal study to explore the development of the timing of turn-taking in infancy. By using fine-grained temporal analyses of vocalizations, we were able to examine the structure and timing of vocal turns and their developmental trajectory in mother–infant exchanges during the

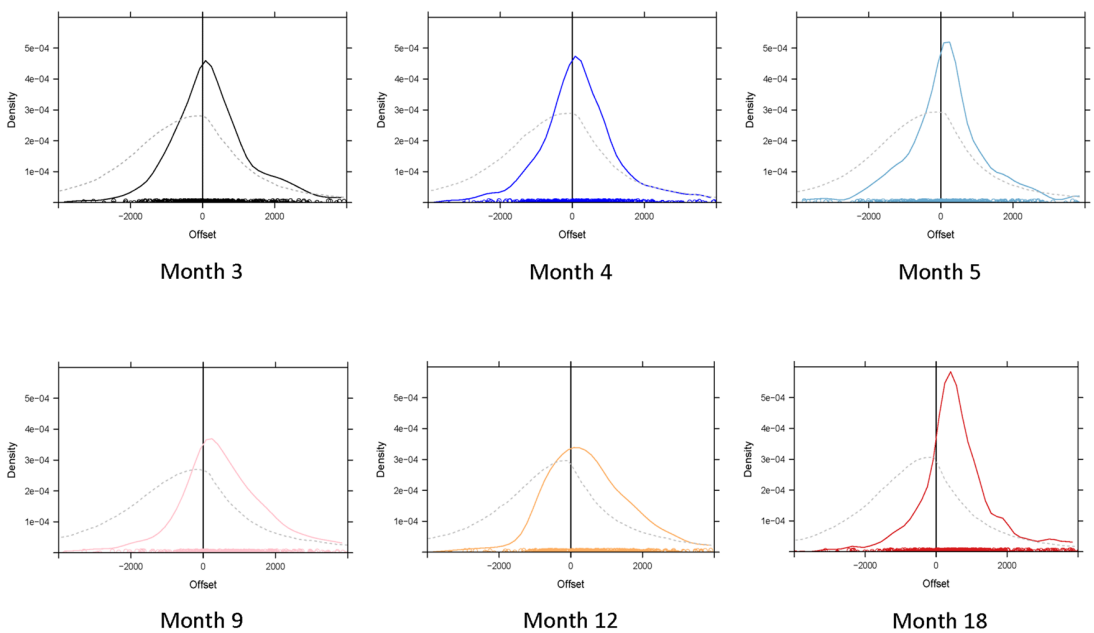


FIGURE 5 | Dashed lines are the random distributions, solid lines the actual distribution.

TABLE 5 | Range of median within-turn pause durations at each age in ms.

	Month 3	Month 4	Month 5	Month 9	Month 12	Month 18
Infant	307–1197.5	320–1208	360–1549	189–5760	420–1925	560–1590
Mother	799–1597	761–1325.5	659–1538	706.5–1559	565–1105	475–1650

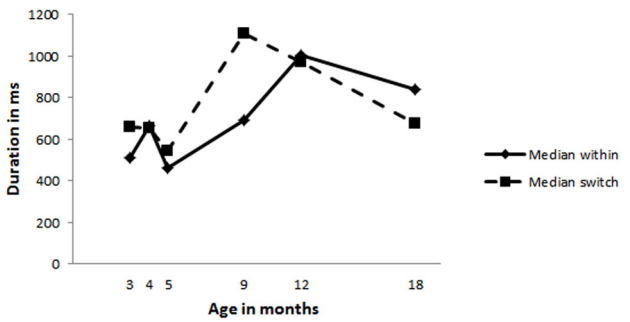


FIGURE 7 | Infants' median within-turn pause durations (solid line) and median gap durations after an maternal vocalization (dashed line).

first year and a half of life. Thus far, research on the timing of turn-taking, or turn-timing, in infancy has been fragmented, mostly cross-sectional, and has either looked at the development of overlapping vocalizations or gap durations but not both (e.g., Bateson, 1975; Jasnow and Feldstein, 1986; Elias and Broerse, 1996). Longitudinal designs including larger samples and tracking development over extended periods of time can provide valuable insights into stability and developmental change in turn-timing skills. We therefore analyzed mother–infant free-play interactions from 3 to 18 months. This allowed us to track the development of turn-timing from early in infancy to an age at which infants begin to acquire productive language. Our aim was to provide a more concise picture of the development of turn-timing, including both the development of overlapping vocalizations and gap durations. Based on previous findings (Garvey and Berninger, 1981; Casillas et al., under review) and the predictions made by the Interaction Engine Hypothesis (Levinson, 2006) we expected that the temporal aspect of turn-taking would develop early in infancy but that, when infants first start to produce language, their timing would slow down. In addition, we explored whether the turn-timing patterns in the mother–infant exchanges were reciprocally structured. As described, we transcribed 10-min free play interactions of 12 mother–infant dyads at the ages of 3, 4, 5, 9, 12, and 18 months, and extracted all turn transitions from these transcriptions. Next we analyzed the turn transitions for the developmental patterns of overlaps, gaps and the transitions as a whole (i.e., both gaps and overlaps).

Overlap

Infants started out by producing just over a third of their turns in overlap with their mothers. This amount remained stable across 3, 4, and 5 months. However, between 5 and 18 months the amount of overlap decreased to about a fifth of their turns. Maternal overlap remained relatively stable over time and at 18 months the percentage of turns produced in overlap by infants was at a similar level to that of their mothers. This finding suggests that from 5 months onward infants adopt more of a turn-taking-like structure in interactions with their mothers. These findings are similar to previous studies reporting a decrease in overlapping vocalizations from 3 to 6 months and onward (Elias and Broerse, 1996) and around 4 months of age onward (Ginsburg and Kilbourne, 1988). In our dataset this decrease in overlapping vocalizations occurred slightly later, which could be due to differences in sample size. For example, Ginsburg and Kilbourne followed three dyads focusing on the period between 2.5 and 5 months. Instead of a decrease, Rutter and Durkin (1987) observed an increase in overlapping vocalizations from 9 to 24 months of age. However, they suggested that this might be due to the increasing amount of vocalizations that infants produced in general. In the present study we analyzed the amount of overlap by calculating the percentage of turns produced in overlap, therefore accounting for differences in total number of turns.

Even though the present data shows that infants start to decrease the amount of overlap that they produce, this does not mean that infants, like adults, aim for the end of the previous turn as the place to launch their own turn (Sacks et al., 1974; Jefferson, 2004). If that were the case we should also see that infants produce shorter overlap durations as they get older. We therefore analyzed the median overlap durations. These analyses revealed that infant overlap durations remained stable with age, while overlap durations of the mothers decreased with infant age. To summarize, mothers and infants start to adopt a more turn-taking like pattern in vocal exchanges from around 5 months onward as evidenced by the decrease in the percentage of overlap that infants produce. However, when infants do produce their turns in overlap they do not yet seem to be aiming, like adults, for the end of the turn.

Gap

Analyses on the median gap durations revealed that the gap durations of the mothers remained fairly stable over time, but infants' gap durations became significantly longer with age. Further exploratory analyses comparing infant gap durations at 5 months with the gap durations at 9, 12, and 18 months respectively revealed that infants were significantly faster at 5 months compared to at 9 and 12 months, but not at 18 months. Together, these findings suggest that infants are initially relatively fast in responding to their mothers' turn, but slow down considerably at 9 months after which they start to pick up speed again. This is consistent with our expectations based on previous work and the Interaction Engine Hypothesis (Garvey and Berninger, 1981; Levinson, 2006; Casillas et al., under review).

However, infants start slowing down somewhat earlier than expected. We expected that the slowing down would coincide with the emergence of productive language, which would be around 12 months instead of at 9 months. Our reasoning was that infants would slow down when they need to integrate these developing linguistic skills with their existing interactional timing skills and 9 months would be somewhat too early for this to occur. An alternative explanation for this finding might be found in infants' changing communicative understanding. The period from 9 months onward is an age at which it has been shown that several skills relevant to communication are emerging, such as joint attention and pointing (e.g., Bakeman and Adamson, 1984; Carpenter et al., 1998). Furthermore, research has shown that this is the period in which infants begin to see others as intentional agents, which is suggested to be the prerequisite for word learning (Carpenter et al., 1998). In other words, infants at this age start to acquire an understanding of the communicative and shared function of social interactions. Thus, the finding that infants are slowing down in their turn-timing around 9 months of age might not be as surprising, as it occurs at an age at which infants are expected to be increasing their communicative understanding of social exchanges. However, more research is necessary to establish whether infants changing communicative understanding is indeed related to the slowing down observed at 9 months. This hypothesis is currently under further exploration. Another explanation might be found in a relation between the decrease in the amount of overlap and the increase in gap durations. It is possible that the decrease in overlap reflects some basic understanding of turn-taking, i.e., waiting to launch your turn until your interlocutor is finished speaking. This in turn could cause the increase in gap duration because infants are waiting until the end of their mother's turn. However, when infants produce overlap at 9 months and older they do not show a decrease in the durations of these overlaps, which is what you would also expect if infants are waiting until the end of the turn. Nevertheless, the possibility that the increase in gap durations is related to the decrease in overlaps deserves further study.

Reciprocity

Work by Anderson et al. (1977), Jasnow and Feldstein (1986), Beebe et al. (1988), Jaffe et al. (2001), showed that infants at 4 months adjusted their gap durations on the basis of who they interacted with. Furthermore, they showed that at 9 months maternal gap durations were influenced by infant gap durations and vice versa. Based on this earlier work we expected to find evidence of reciprocally structured turn-timing patterns at all ages. This is exactly what we found. In the present study maternal turn-timing remains fairly stable over time, while infant turn-timing is changing considerably: the amount of overlapping vocalizations decreases with infant age, while their gap durations seem to be much longer around 9 and 12 months of age compared to the gap durations at 3, 4, and 5 months. But, even though this indicates that infants are changing their turn-timing behavior over time, it is still possible that infants were vocalizing randomly while mothers were mainly responsible for establishing a turn-taking structure. Mothers could have changed, for example, how long they will wait for their infant to produce a turn after their own utterances, i.e., the duration of mothers' within-turn pause. Longer maternal within-turn pause durations could explain a decrease in infant overlapping vocalizations because infants are given more time to respond. We therefore ran two analyses to assess the reciprocity of the vocal exchanges. We first analyzed whether the observed distribution of median durations of turn transitions was significantly different compared to randomly sampled distributions. This analysis revealed that the observed data was different from random distributions at all ages. Next we analyzed maternal within-turn pause durations. This revealed that mothers do not change their within-turn pause durations with increasing infant age. They wait equally long for an infant turn regardless of infant age. Thus the decrease in the percentage of turns that infants produce in overlap cannot be explained by a change in maternal within-turn pauses. Mothers do produce longer within-turn pauses compared to their gap durations when responding to an infant vocalization, which is similar to what Bateson (1975) found in her observation of mother–infant vocal exchanges. This suggests that in general, at all observed ages, mothers respond faster after an infant vocalization compared to after their own utterances. Bateson also observed a similar trend for the infant in her study, which is different from what we observed in the present data. This difference could be due to differences in definitions: Bateson calculated the within-turn pause from the onset of the utterance until the onset of the next utterance, i.e., including the preceding utterance, while in the present study within-turn pauses were calculated from the end of utterance until the beginning of the next utterance. In addition, there is also a difference in sample size: Bateson followed one dyad from 1.5 to 3.5 months while the present study followed 12 mother–infant dyads across a longer period of time. Our findings show clear changes in infant turn-timing skills which do not seem to be due to differences in maternal turn-timing as mothers remain stable with regards to the amount of overlap they produce, their gap durations and their within-turn pause durations. Infants thus seem to actively contribute to the observed changes in turn-timing.

The present findings provide some initial support for the Interaction Engine hypothesis, especially the findings with regards to the gap durations which are relatively short early in infancy but have increased considerably around 9 months. This slowing down coincides with a period of important changes in infants' communicative and social understanding of interactions (e.g., Bakeman and Adamson, 1984; Carpenter et al., 1998). However, more research is needed to further explore what exactly is driving this change in turn-timing and whether this slowing down is related to infants' advancing communicative understanding of interaction. Future studies should combine experimental methods that assess infants' turn-timing skills with assessments (e.g., experiments or parental report) of infants' language comprehension and production skills, to disentangle possible links between infant turn-timing and language production and between infant turn-timing and their understanding of language. Moreover, research should aim to assess whether the changes observed in timing in this study are

related to infants' changing understanding of the communicative and social function of social exchanges. One possible explanation for why infants are slowing down we are currently exploring, is whether the complexity of infant vocalizations might be related to turn-timing. A recent study by Casillas et al. (under review) has demonstrated such complexity effects on timing in older children. Furthermore, the present study solely focused on turn-timing in vocal exchanges, while turn-taking, and thus turn-timing, occurs from early on in infancy in various types of social exchanges, not necessarily just in vocal exchanges. For example, give and take sequences involving objects occur from around 9 months of age onward (e.g., Bates et al., 1975). Another option would be to look at the timing of pointing which emerges between 9 and 12 months. It could be, for example, that infants respond faster when pointing compared to when using a vocal response. Analyses of the timing involved in these types of turn-taking sequences might give additional insights into the role of timing in interaction in general versus timing specifically related to vocal exchanges.

The present study found no indication that changes in maternal timing were responsible for the changes in infant timing. However, mothers might have been changing other behaviors that influenced infant timing, including the use of gestures, facial expressions or changes in the content of the exchanges. For example, a mother could lean forward toward her child as to indicate 'I am handing you the turn,' which could facilitate turn-timing. But, if such cues were influencing the vocal-timing assessed in the present study one would expect that infants are becoming better with age at interpreting these cues and therefore will speed up with age instead of slowing down. Nevertheless, future studies should address the use and role of multimodal cues on turn-timing. The use of motion sensors could allow for conducting analyses at the same fine-grained level as with vocal turn-timing in the present paper. In addition, questions remain on how much of the changes in infant timing are due to social interactional experience. Based on studies assessing short-term effects of contingent and non-contingent interaction on infant behavior it seems likely that early interactional experience plays an important role (e.g., Bloom et al., 1987; Bloom, 1988; Masataka, 1993). Nevertheless, more research is needed to further explore the role of social interactional experience in the first few months of life on the development of infant turn-timing skills. For example, short training studies in which parents are trained to provide contingent feedback could provide insights into the impact of contingent experience in interaction on infant turn-timing skills (e.g., McGillion et al., 2014). In addition, studies including different types of samples, such as infants of mothers who suffer from postnatal depression, can also shed light on the role of interactional experience (e.g., Field, 1984). Finally, research should not ignore the infants' possible role in the development of early interactional skills. Infants are likely to differ in how many opportunities they provide for their mothers to respond, by gazing, smiling, and vocalizing at their mother. Thus, individual differences in infant characteristics could also play a role in infants' interactional experiences and the development of infant turn-timing.

The present study is, to our knowledge, the first study to assess turn-timing in infancy including both overlap and gap. In addition, we believe the present study is the first to provide a comprehensive overview of this development including not only analyses on the amount of overlapping vocalizations, but also assess the duration of overlap. The longitudinal design of the study has allowed us to demonstrate that infants' turn-timing skills are changing considerably during infancy and that these changes occur around the same time as when infants' communicative understanding has been found to be changing (e.g., Bakeman and Adamson, 1984; Carpenter et al., 1998). Furthermore, maternal turn-timing does not change much over this period of time indicating that the infants are actively involved in this observed developmental change. The observed developmental pattern is consistent with earlier research (e.g., Garvey and Berninger, 1981; Ginsburg and Kilbourne, 1988) and the predictions of the Interaction Engine Hypothesis (Levinson, 2006). Finally, the finding that infants are relatively fast turn-timers at 3, 4, and 5 months highlights the existence of remarkable social interactional abilities early in infancy.

Acknowledgments

This research was supported by the ERC Advanced Grant #269484 INTERACT awarded to SL, the Leverhulme Trust and the Economic and Social Sciences Research Council. We thank the First Steps team: Elena Sakkalou, Kate Ellis-Davies, and Nia Fowler for their contributions to the data collection, Sean Roberts for helping with the analyses, Paula Fikkert, Imme Lammertink, Marisa Casillas, and Connie de Vos for their comments, and the Language and Cognition Department.

References

Anderson, B. J., Vietze, P., and Dokecki, P. R. (1977). Reciprocity in vocal interactions of mothers and infants. *Child Dev.* 48, 1676–1681. doi: 10.2307/1128534

Bakeman, R., and Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Dev.* 55, 1278–1289. doi: 10.2307/1129997

Bates, E., Camaioni, L., and Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Q.* 21, 205–226.

Bates, D., Maechler, M., and Dai, B. (2012). *Lme4: Linear Mixedeffects Models Using S4 Classes. [Computer Software Manual]*. Available at: http://lme4.r-forge.rproject.org/

Bateson, M. C. (1975). Mother-infant exchanges: the epigenesis of conversational interaction. *Ann. N. Y. Acad. Sci.* 263, 101–113. doi: 10.1111/j.1749-6632.1975.tb41575.x

Beebe, B., Alson, D., Jaffe, J., Feldstein, S., and Crown, C. (1988). Vocal congruence in mother-infant play. *J. Psycholinguist. Res.* 17, 245–259. doi: 10.1007/BF01686358

Bloom, K. (1988). Quality of adult vocalizations affects the quality of infant vocalizations. *J. Child Lang.* 15, 469–480. doi: 10.1017/S0305000900012502

Bloom, K., Russell, A., and Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *J. Child Lang.* 14, 211–227. doi: 10.1017/S0305000900012897

Bruner, J. S. (1975). The ontogenesis of speech acts. *J. Child Lang.* 2, 1–19. doi: 10.1017/S0305000900000866

Bruner, J. S. (1981). The social context of language acquisition. *Lang. Commun.* 1, 155–178. doi: 10.1016/0271-5309(81)90010-0

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63, 1–174. doi: 10.2307/1166214

Casillas, M., and Frank, M. C. (2013). "The development of predictive processes in children's discourse understanding," in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Society), 299–304.

Elias, G., and Broerse, J. (1996). Developmental changes in the incidence and likelihood of simultaneous talk during the first two years: a question of function. *J. Child Lang.* 23, 201–217. doi: 10.1017/S0305000900010151

Elias, G., Hayes, A., and Broerse, J. (1986). Maternal control of co-vocalization and inter-speaker silences in mother-infant vocal engagements. *J. Child Psychol. Psychiatry* 27, 409–415. doi: 10.1111/j.1469-7610.1986.tb01842.x

Ellis-Davies, K., Sakkalou, E., Fowler, N. C., Hilbrink, E. E., and Gattis, M. (2012). CUE: the continuous unified electronic diary method. *Behav. Res. Methods* 44, 1063–1078. doi: 10.3758/s13428-012-0205-1

Field, T. (1984). Early interactions between infants and their postpartum depressed mothers. *Infant Behavior. Dev.* 7, 527–532. doi: 10.1016/S0163-6383(84)80012-0

Garvey, C., and Berninger, G. (1981). Timing and turn taking in children's conversations. *Discourse Process.* 4, 27–57. doi: 10.1080/01638538109544505

Ginsburg, G. P., and Kilbourne, B. (1988). Emergence of vocal alternation in mother-infant interchanges. *J. Child Lang.* 15, 221–235. doi: 10.1017/S0305000900012344

Goldstein, M. H., King, A. P., and West, M. J. (2003). Social interaction shapes babbling: testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8030–8035. doi: 10.1073/pnas.1332441100

Hilbrink, E., Sakkalou, E., Ellis-Davies, K., Fowler, N., and Gattis, M. (2013). Selective and faithful imitation at 12 and 15 months. *Dev. Sci.* 16, 828–840. doi: 10.1111/desc.12070

Hoicka, E. K., and Akhtar, N. (2011). Preschoolers joke with jokers, but correct foreigners. *Dev. Sci.* 14, 848–858. doi: 10.1111/j.1467-7687.2010.01033.x

Hsu, H., and Fogel, A. (2003). Stability and transitions in mother-infant face-to-face communication during the first 6 months: a microhistorical approach. *Dev. Psychol.* 39, 1061–1082. doi: 10.1037/0012-1649.39.6.1061

Hsu, H., Fogel, A., and Messinger, D. S. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: factors associated with quantitative and qualitative differences. *Infant Behav. Dev.* 24, 107–128. doi: 10.1016/S0163-6383(01)00061-3

Jaffe, J., Beebe, B., Feldstein, S., Crown, C., and Jasnow, M. D. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monogr. Soc. Res. Child Dev.* 66, 1–8.

Jasnow, M., and Feldstein, S. (1986). Adult-like temporal characteristics of mother-infant vocal interactions. *Child Dev.* 57, 754–761. doi: 10.2307/1130352

Jefferson, G. (2004). "A sketch of some orderly aspects of overlap in natural conversation," in *Conversation Analysis. Studies from the First Generation*, ed. G. H. Lerner (Amsterdam: John Benjamins Publishing Company), 43–59.

Kaye, K. (1977). "Toward the origin of dialogue," in *Studies in Mother-Infant Interaction*, ed. H. R. Schaffer (London: Academic Press), 89–117.

Kaye, K., and Fogel, A. (1980). The temporal structure of face-to-face communication between mother and infants. *Dev. Psychol.* 16, 454–464. doi: 10.1037/0012-1649.16.5.454

Levinson, S. C. (2006). "On the human "interaction engine"," in *Roots of Human Sociality. Culture, Cognition and Interaction*, eds N. J. Enfield and S. C. Levinson (New York, NY: Berg), 39–69.

Levinson, S. C. (2013). "Action formation and ascription," in *The Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Malden, MA: Wiley-Blackwell), 103–130.

Masataka, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behavior of three- to four-month-old Japanese infants. *J. Child Lang.* 20, 303–312. doi: 10.1017/S0305000900008291

McGillion, M., Pine, J., Herbert, J., and Matthews, D. (2014). *A Training Study to Promote Contingent Talk and Test its Effect on Language Development*. Amsterdam: Talk at IASCL.

Murray, L., and Trevarthen, C. (1985). "Emotional regulation of interactions between two-month-olds and their mothers," in *Social Perception in Infants*, eds T. M. Field and N. A. Fox (Norwood, NJ: Alex Publishing Corporation), 177–197.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/

Ratner, N., and Bruner, J. (1978). Games, social exchanges and the acquisition of language. *J. Child Lang.* 5, 391–401. doi: 10.1017/S0305000900002063

Rochat, P., Querido, J. G., and Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Dev. Psychol.* 35, 950–957. doi: 10.1037/0012-1649.35.4.950

Rosenthal, M. K. (1982). Vocal dialogues in the neonatal period. *Dev. Psychol.* 18, 17–21. doi: 10.1037/0012-1649.18.1.17

Rutter, D. R., and Durkin, K. (1987). Turn-taking in mother-infant interaction: an examination of vocalizations and gaze. *Dev. Psychol.* 23, 54–61. doi: 10.1037/0012-1649.23.1.54

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.2307/412243

Sloetjes, H., and Wittenburg, P. (2008). "Annotation by category – ELAN, and ISO DCR," in *Proceedings of the Sixth International Conference on Language Resources and Evaluation* (Marrakech: LREC 2008).

Snow, C. E. (1977). The development of conversation between mothers and babies. *J. Child Lang.* 4, 1–22. doi: 10.1017/S0305000900000453

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Striano, T., Henning, A., and Stahl, D. (2006). Sensitivity to interpersonal timing at 3 and 6 months of age. *Interact. Stud.* 7, 251–271. doi: 10.1075/is.7.2.08str

ten Bosch, L., Oostdijk, N., and Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Commun.* 47, 80–86. doi: 10.1016/j.specom.2005.05.009

Thorgrimsson, G. (2014). *Infants' Understanding of Communication as Participants and Observers*. Ph.D. thesis, Radboud University Nijmegen, Nijmegen.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 July 2015
doi: 10.3389/fpsyg.2015.00890

Turn-taking: a case study of early gesture and word use in answering WHERE and WHICH questions

Eve V. Clark and Kate L. Lindsey*

Department of Linguistics, Stanford University, Stanford, CA, USA

When young children answer questions, they do so more slowly than adults and appear to have difficulty finding the appropriate words. Because children leave gaps before they respond, it is possible that they could answer faster with gestures than with words. In this study, we compare gestural and verbal responses from one child between the ages of 1;4 and 3;5, to adult **Where** and **Which** questions, which can be answered with gestures and/or words. After extracting all adult **Where** and **Which** questions and child answers from longitudinal videotaped sessions, we examined the timing from the end of each question to the start of the response, and compared the timing for gestures and words. Child responses could take the form of a gesture or word(s); the latter could be words repeated from the adult question or new words retrieved by the child. Or responses could be complex: a gesture + word repeat, gesture + new word, or word repeat + new word. Gestures were the fastest overall, followed successively by word-repeats, then new-word responses. This ordering, with gestures ahead of words, suggests that the child knows what to answer but needs more time to retrieve any relevant words. In short, word retrieval and articulation appear to be bottlenecks in the timing of responses: both add to the planning required in answering a question.

Keywords: where and which questions, answers, gestures, words, timing

OPEN ACCESS

Edited by:
Kobin H. Kendrick, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Elma Hilbrink, Max Planck Institute for Psycholinguistics, Netherlands
Thea Cameron-Faulkner, University of Manchester, UK

****Correspondence:***
Eve V. Clark, Department of Linguistics, Stanford University, Margaret Jacks Hall, Building 460, Stanford, CA 94305-2150, USA
eclark@stanford.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *03 November 2014*
Accepted: *15 June 2015*
Published: *08 July 2015*

Citation:
Clark EV and Lindsey KL (2015) Turn-taking: a case study of early gesture and word use in answering WHERE and WHICH questions. Front. Psychol. 6:890. doi: 10.3389/fpsyg.2015.00890

Introduction

When adults answer questions, their answers are surprisingly fast, regardless of the language involved. The median gap between the end of a yes/no question, for example, and the start of the answer, is about 200 ms for speakers of English, with a range over other languages from 0 ms to nearly 400 ms (Stivers et al., 2009). This close timing entails that the addressee begins planning an answer before the speaker finishes the question. This timing reflects the general principle of 'no gap, no overlap' as speakers participate in conversation.

Young children take much longer to answer questions than adults do (Casillas, 2014a), apparently waiting until they have heard the full question before they start to plan an answer. To interpret a question, children must first process its content. [Between 1;6 (1 year, 6 months) and 3;0, they get much faster, almost to adult speed, in recognizing familiar words (e.g., Zangl et al., 2005; Fernald et al., 2006)]. Once they have done this, they can then decide on and formulate an answer. But they may take up to 2 s, or more, before they start their answer after a question. The size of this gap between question and answer decreases over time for specific question-types (Casillas et al., under review), and by around 3;6–4;0, children often manage near-adult timing.

To date, studies of how children answer questions have focussed on verbal answers, where children either repeat one or more words from the adult question in their answer, or construct a verbal answer from scratch. This suggests that one factor in answering a question is how much word production children have to plan in order to come up with an answer. When they repeat a word from the adult's question as the answer, they need not retrieve anything from memory, so the cost should be minimal. But when they have to retrieve one or more appropriate words, and the relevant construction, from memory before they begin to articulate their answer, the cost should be much greater.

What happens, though, when children (or adults, come to that), have the option of answering the question with a gesture, either alone or in combination with speech? Once children have decided on an answer, to extend hand and arm to designate a place in answer to a *Where* question, or to select an alternative in answer to a *Which* question, would seem to require less planning and articulation than producing an utterance, where children must first decide which words to repeat of those already said, say, or else search for and retrieve one or more appropriate words and then articulate those in their response.

Gestures in Young Children

Children start to produce manual gestures fairly early to indicate interest in or desire for something in their immediate environment. They start to point (interest: 'look') and reach (desire: 'I want') from as early as 10–12 months. These two gesture-types have been considered to be proto-speech acts for asserting and requesting (Werner and Kaplan, 1963; Wundt, 1973; Bates et al., 1975; Bruner, 1975). The vast majority of young children's early gestures appear to be deictic in nature, mainly index-finger POINTS, but they also produce gestures to SHOW, OFFER, or PLACE objects for the other (see Caselli, 1990; Liszkowski, 2006; Andrén, 2010). In his case study of five children up to age 2;6, Andrén (2010) found that these gestures were nearly always (94%) coordinated with speech, with only a few gestures overall occurring on their own (see also Kelly, 2014).

Children demonstrate comprehension as well as production of POINTS from around 12 months on (e.g., Muñetón Ayala and Rodrigo López, 2011; Behne et al., 2012). When adults look and point, children treat this as a directive to look at whatever is being targeted. In general, adult gaze and pointing are critical in establishing joint attention (e.g., Carpenter et al., 1998; Estigarribia and Clark, 2007).

By 1;6, children appear to use POINTS differentially: to elicit a term for something, they simply point and look, and adults then typically offer a label for that referent. But when children combine a point with vocalization or a word, adults are significantly more likely to treat that combination as a request and to give the child the object targeted (see Carter, 1978; Kelly, 2011; Olson and Masur, 2011).

Where and *Which* Questions

When young children answer questions, they may substitute an indicating gesture for speech, for example answering a *where* question with a POINT to the relevant location. In this study, we compare gestural and verbal responses to adult *Where* and *Which* questions addressed to one child between the ages 1;4 and 3;5.

To make this comparison, we examined the relative timing of any gesture onset in an answer, compared to the timing of any vocal or verbal onset in the answer. Answers could take several forms: (a) they could be simple, containing just a single response-element: a gesture, a babble (vocalization), or a word—either a word repeated from the adult question or a new verbal choice made by the child; or (b) they could be complex, consisting of two elements: a gesture + babble, gesture + repeat, gesture + new-word(s), or repeat + new-word(s).

Since gestural responses do not require any lexical retrieval, we expect that children should be able to produce a gestural response as soon as they have understood the question. Producing a new verbal response should be more complex and therefore take more time: the child must retrieve any relevant word(s), plan the answer, and then articulate it. So any measurable difference between gestural and verbal response times would provide a preliminary indication of question-comprehension time (how long it takes to issue a gesture) and hence how much time is required to produce a verbal response.

If the child produced a gesture in a combined gesture + verbal response, with the gesture produced ahead of the child's word(s), this would be evidence that the child knows what to answer but needs additional time to retrieve relevant words. The conceptual cost of identifying a relevant answer is the same for gestural and verbal responses. But for verbal responses, lexical retrieval is one bottleneck in the timing of an answer, followed by planning the response and articulation (Levelt, 1989; Levinson, 2000, 2006). We would therefore expect that repeats of one or more words from the adult question, available in short-term memory, should take less time than when the child constructs an answer with words newly retrieved from memory. In short, both lexical retrieval and articulation are added costs and so should take longer, as shown schematically in **Table 1**. In the present case study we focus on the processing costs in the production of verbal responses compared to gestural ones.

Materials and Methods

For this study, we drew on the corpus for one child, Alex, from the Providence Corpus of American English (Demuth et al., 2006), in the Child Language Data Exchange System archive (CHILDES; MacWhinney, 2000). All data in the CHILDES Archive were collected in accordance with the internal review board on human subjects of the relevant university, with permission for use of the data in further analyses by researchers not involved

TABLE 1 | Costs (effort) required in answering a question.

	Lexical retrieval	Articulation
Gesture	–	–
Babble	–	(√)
Repeated word(s)	–	√
New word(s)	√	√

in the original study. This corpus contains high-quality video recordings of biweekly spontaneous interactions, each lasting 1–1.5 h, between parent and child. Each video session contains numerous parental questions and child answers. This allowed us to measure both gestural and verbal onset times in the child's responses to parental questions. Recordings for Alex were made approximately every 2 weeks, beginning at age 1;4.20 until 3;5.15, for a total of 51 video sessions. For our analysis, we chose 26 of these sessions, one per month, to capture snapshots of Alex's development. If any particular month had fewer than 10 tokens of the relevant question–answer types, and if there was an additional video available in that month, we drew on both video sessions. We did this for nine sessions—at 1;7, 1;11, 2;2, 2;3, 2;8, 2;10, 3;1, 3;2, and 3;4—so each of these months was represented by two data sessions from Alex, for a total of 35 videos in all.

In order to analyze the time it took for the child to answer his mother's questions, we compared gestural and verbal responses. Gestural responses offer a non-verbal form of response that avoids the need to find appropriate words, while verbal responses require finding and producing an appropriate answer. *Where* and *Which* questions can be answered with either a gestural or verbal response, or both. We extracted all *Where* and *Which* questions, and measured the onset times of all responses provided by the child. We identified a total of 502 *Where* and *Which* questions in the 35 videos selected, and then identified all the responses given. The adult (mother) in each case treated the child's responses as answers to the question just asked. Moreover, the majority of these responses explicitly provided relevant semantic information: a target-place in response to *Where* questions, and a chosen alternative in response to *Which* questions. The one response-type we were unable to assess was vocal babble on its own: these babbles were consistent in form but, to us, uninterpretable. His mother, however, treated these too as appropriate responses.

Coding

We coded all the child responses using the following categories: (i) *gesture*: manual gesture (G) to the target object's location, (ii) *speech*: babble (B), repeat of a term from the adult question (R), new verbal response (V), and (iii) *location*: on camera or off camera (O). We also coded as 'no response' (N) any questions the child failed to answer, questions where the child was impeded from responding (i.e., while eating, or with a body position that delayed a possible gestural response), and questions asked when the child was already manipulating, or speaking about, the relevant object.

We collected metadata for each question/response pair, including: (i) age of child, (ii) video file, (iii) timestamp for that question in the video session (HH:MM:SS), (iv) question type (*Where* or *Which*), (v) the adult's actual question, and (vi) the content of the child's response. Next, we used a Python script to extract 12 s of audio/video for each of the 502 questions we had identified, beginning 2 s before each question onset. Each video clip was imported into ELAN and paired with a synchronized transcription file. The onsets and endings of all questions, and of all answers (gestural and verbal) were transcribed into question, gestural, and verbal tiers, respectively. Adult *Where* and *Which* questions received 235 (47%) unimpeded gestural and/or verbal responses from the child.

We then measured, in milliseconds, the time from the ending of each question to the onset of the answer. Question endings were marked following complete expulsion of sound, including aspiration. Verbal onsets were marked immediately prior to any vocal utterance (including utterance-initial "um"). Gesture onsets were marked immediately prior to the beginning of movement (the preparation of the gesture), and duration of the stroke was also measured where possible. In the nine instances where a question was repeated multiple times, we used our discretion in determining which instance the child was attending to. Lastly, we excluded outlier responses that were two SDs or more from the relevant mean (13 answers, 5.5%) from our statistical analyses.

The timing of 10% of Alex's responses to *Where* and *Which* questions was rechecked by an independent coder. This coder and the second author agreed on the timing measurements, within 200 ms (the smallest discriminable difference) 88% of the time, with high inter-rater reliability (Pearson's $r = 0.930$, $p < 0.0001$). When the timing window was expanded to within 400 ms, the agreement rate rose only very slightly, to 91%.

Results

Of these 221 questions included in our statistical analyses, *Where* questions received 137 responses, and *Which* questions 84 responses. 'Single', simple, responses followed 145 (66%), and 'double', complex, responses followed 76 (34%). In the double responses, we measured the onset timing for each element separately (76×2). The single or simple responses consisted of one element in the response. There were 14 gestures, 13 babbles, 56 repeats, and 62 new-word responses (i.e., words retrieved specifically for the response given), as shown in **Table 3**. Double or complex responses consisted of those where the child combined a gesture with a vocalization or word, or combined a repeat from the adult's question with some added verbal material. These combinations were ordered, with 10 gesture + babble responses, 7 gesture + repeated word, 39 gesture + new-word, and 2 repeated word + new-word responses, along with another 18 other complex answers where the second element seemed to be produced for added clarification rather than being an integral part of the initial response. In these cases, Alex produced an indicating gesture *after* saying a repeated or new word. In seven instances, he babbled and then, after a pause, pointed; in five more, he repeated a word from the question, and only then pointed; and in another five, he produced a new word and then pointed. In each instance, Alex appeared to be trying to clarify his answer by adding the gesture afterward (see **Table 4**, right-hand columns). And in one, he produced a one-word answer, paused, and then repeated a word from the adult question. We therefore excluded the timing for these from our overall computations of the relative timing of the two parts in double responses.

In all, we measured 297 response tokens (145 tokens from the simple responses, and 152 tokens from the 76 complex, double

responses). The overall mean response times, from the end of the question to the onset of the child's answer, are summarized in **Table 2**.

Overall, Alex's gestures were produced faster than his babble responses. His babbles, repeats of adult words, and new-word responses differed significantly overall [$F(3,259) = 9.355$, $p < 0.0001$], with babbles and repeats both produced significantly faster than new-word responses (**Table 2**). His gestures were faster than both repeated word responses (Fisher's Least Significant Difference test, $p < 0.021$), and new-word responses (Fisher's LSD, $p < 0.0001$). And his repeated word responses in turn were significantly faster than his new-word responses (Fisher's LSD, $p < 0.009$). Finally, within his complex double responses (**Table 4**), his gestures were significantly faster than the verbal responses they were paired with, where the gesture onset preceded the word onset [$t(56) = 7.408$, $p < 0.0001$].

Although we combined the responses to *Where* and *Which* questions in the overall analysis, inspection of the two questions and their responses separately showed the same patterns in onset timing for the response types: Gestures (*Where* 598 ms, *Which* 379 ms) were produced faster than any verbal responses. Among verbal responses, babbles and repeats (*Where* 1056, 753; *Which* 580, 353) were faster than new-word responses (*Where* 1103; *Which* 848), and repeats in turn were faster than new-word responses.

When we compare these response-types in single responses (**Table 3**) to double responses (**Table 4**), we see that the mean onset timing for gestures and verbal responses varies with the complexity of the answer. It also varies for verbal responses, whether the child repeats one or more of the words he has just heard, combines one or more repeated words with a word newly retrieved from memory, or constructs a verbal response of one to three or more words that are entirely new. Although responses to some question-types get faster with age (Casillas et al., under review), such gains are often obscured by the fact that, as children get older, they also start to produce some longer, more complex answers where earlier they had produced only one word.

The single gesture responses consistently indicated the location in response to *Where*, and the alternative-chosen in response to *Which*. Babbles, produced on their own and in some double responses, may have had some attention-getting function, especially when combined with a pointing gesture. On their own, they appear to have been interpretable to the mother as early attempts at words: she treated them all as responses. The verbal elements in single responses consisted either of one or more words repeated from the adult question ($n = 56$)—typically the word for the object being sought (*Where?*) or for the object chosen (*Which?*)—or of semantically relevant unmentioned words retrieved by the child from memory ($n = 62$). Initially, Alex tended to repeat words, mainly single words, from the adult question (e.g., *papa, balloon*, etc.), but as he got older, he also began to produce deictic terms like *there* or *that*, either on their own or in combination with repeated or newly retrieved words. His verbal responses also became longer with age, and by 2;4–2;6, he produced some answers of three words or more [e.g., Which shoes, the green ones? —*The green ones* (2;6); Where would you like to sit, right here? —*Over there* (2;8); Where's the baby going? —*Baby go in a stroller to go for a walk* (3;0); Which one would you like to paint? —*I would like to paint this one* (3;1)].

Does planning a longer verbal response take longer? We looked at the correlations with age for Alex's responses (a) where he repeated one vs. two vs. three or more words from the adult question, and (b) where he produced one vs. two vs. three or more words retrieved from memory. In the first case, where he had just heard the word(s) he repeated in the adult's question, there was no change with age, whether he repeated 1, 2, or 3+ words ($r = 0.085$, n.s.), but when Alex constructed his answer with his own words, he produced a larger number of more complex (longer) utterances as he got older ($r = 0.225$, $p < 0.043$). We also looked, within his single responses, at how long he took to produce 1–, 2–, and 3+ word answers. On average, longer answers took longer to produce for both repeats and new-word responses, as shown in **Table 5**. However, there was effectively no correlation between length-of-response and timing here, largely because both repeats ($r = 0.108$, n.s.) and new words ($r = 0.149$, $p < 0.065$, n.s.) displayed wide variance in timing with production of utterances of the same length. This is attributable to extraneous factors such as how well the child was actually attending when the adult issued the question, whether he could remember immediately where something was, and his basis for deciding which alternative to choose.

As children get older, they get better at planning and better at articulating words and sequences of words, so they can produce longer answers. But they should also gradually speed up, perhaps doing so more readily when they repeat words from the adult question than when they construct an entirely new answer themselves. This would initially yield the different means for responses where Alex repeated one or more words from the adult question, compared to where he constructed a new answer with words retrieved from memory (see **Table 5**). As Casillas et al. (under review) noted, young children do get faster at answering *yes/no* questions as they get older. However, children's increasing speed in answering *Wh*-questions is often obscured by their ongoing acquisition of different *Wh*-words, and the added planning needed to answer *Wh*-questions, as

TABLE 2 | Mean response times (in milliseconds) for all response-types.

Response-type	*N*	Time to response onset
Gesture	70	504
Babble	23	725
Repeat	65	717
New word	103	957

TABLE 3 | Mean response times (in milliseconds) for single response-types.

Response-type	N	Time to response onset
Gesture	14	788
Babble	13	685
Repeat	56	678
New-word	62	950

TABLE 4 | Mean response times (in milliseconds) for double response-types, with the basic order of the two elements produced.

Response-type	*N*	Time-1	Time-2	Response-type	*N*	Time-1	Time-2
Gesture + Babble	10	258	778	Babble + Gesture	5	553	678
Gesture + Repeat	7	567	916	Repeat + Gesture	7	805	984
Gesture + New-word	39	454	919	New-word + Gesture	5	784	1422
Repeat + New-word	2	1101	1919	New-word + Repeat	1	546	1209

children master the meaning of each *Wh*–type in turn. Of *Where* and *Which*, *Where* is typically acquired first (Ervin-Tripp, 1970; Tyack and Ingram, 1977), and Alex answered *Where* questions almost twice as often as *Which* (137 to 84). However, we did not have enough data to detect any changes in speed for Alex's answers to these *Wh*- questions over time, as he got older.

Discussion

We analyzed the response times for all answer-types and, as expected, found that Alex's gestures were produced the fastest overall. This finding is consistent with our expectation that both word-retrieval and articulation add costs to responding. Young children struggle to retrieve words from memory, and also have a hard time producing words in recognizable form. When we limited our analysis to his verbal responses, we found that Alex took less time overall when he relied on one or more words repeated from the adult's question than verbal responses when he retrieved and produced specific words of his own. Here too, we see a cost for word-retrieval, as against simply repeating words available in short term memory. In double answers, Alex generally produced gestures ahead of words, suggesting that he typically knew what he wanted to answer, but needed time both to retrieve words and to articulate them. In short, word retrieval is one bottleneck in the timing of responses: it adds to the planning cost required in answering a question.

Recognizing and Retrieving Words

Children get better at recognizing familiar words as they get older. They speed up steadily from 15 to 24 months, at which time they come close to achieving adult speed in recognizing familiar words (see Fernald et al., 2006, 2013; Fernald and Marchman, 2012). They also steadily improve in recognizing partial words, and in processing words that have been mispronounced (e.g., Swingley and Aslin, 2000; Swingley, 2009).

Recognizing familiar words, though, is a rather different matter from retrieving those words and producing them in order to answer a question. During their second year, as children's production vocabularies begin to expand, they make numerous errors in production, often retrieving the wrong word. Dapretto and Bjork (2000) found that children between 14 and 24 months with larger vocabularies were more likely to be able to retrieve the appropriate words for objects that had been hidden in a box, and that with pictures as prompts they could generally retrieve the appropriate words. They also found that retrieval errors were very frequent in naturalistic picture-book reading, for those children whose production vocabulary had just begun to increase, compared to those who still had only a very small vocabulary or those with a relatively large vocabulary toward the end of the second year. Retrieving the right words early on, then, gives young children much more difficulty than recognizing words they hear from others. This is consistent with the general advance of comprehension over production (see Clark and Hecht, 1983).

But in order to answer questions, children need to be able to retrieve the right words. This in turn depends on their having already made the appropriate mapping as they linked forms and (preliminary) word meanings so they could recognize those words from others—for comprehension. This first step is essential for children trying to retrieve the relevant or most appropriate word(s) for production. Question-answering depends on both comprehension and production, with comprehension of the question followed up by (a) the idea of a possible answer, and (b) its instantiation as a gesture, as an utterance, or as some combination of the two (see **Tables 3** and **4**). But responding with an utterance requires that children be able to retrieve any pertinent words and, if necessary, combine two or more words in a syntactic construction, for subsequent articulation.

What Role does Articulation Play?

In learning how to produce a word, children need to produce it in a form that is recognizable to the addressee, but that may take quite a long time to achieve. Children's early attempts at words often fail because they do not produce a recognizable word and because they produce different variants each time they try to say that word (see, e.g., Dromi, 1987). This, of course, makes it harder for adults to recognize what the child is trying to say. At the same time, if children produce *consistent* word forms, as when children rely on idiosyncratic templates (e.g., /babiŋk/ for 'blanket'), even if these fail to match the adult targets, adults can generally understand what the children intend. But children continually monitor and fine-tune their production, eventually matching the conventional forms produced in the speech community around them. One way to characterize their articulatory development is in the form of two 'rules' for early

TABLE 5 | Alex's response length in words and timing in milliseconds.

Repeats			New words		
Length	***N***	**msecs**	**Length**	***N***	**msecs**
1 word	41	669	1 word	38	867
2 words	16	752	2 words	39	877
3+ words	6	818	3+ words	28	1109

word production, *Be consistent* and *Be precise* (McAllister Byun et al., 2016). The former allows for recognition by others, across different contexts, of words articulated in a non-standard way, and the latter captures the fact that children try hard to emulate the conventional forms they hear from adults. Learning how to articulate single words and then sequences of two or more words takes effort, and acquiring the adult pronunciations takes time (Stoel-Gammon, 2011). In short, articulating any kind of verbal response to a question takes time, over and above the time needed to retrieve words from memory. Planning a response, from deciding on an appropriate conceptual answer, to finding the words and organizing them into an appropriate construction, to articulating the relevant utterance, all takes time (Levelt, 1989; see also Levinson, 2006).

Finally, in relying on gestures alone and on gestures along with some speech, children are following adult usage. When adults take turns, and, for example, ask and answer questions, they can do this using speech or gesture, or both (Clark, 2012). And gestures are often used in interaction to identify the place of some target referent, along with a description, or in lieu of a description (e.g., Bangerter and Oppenheimer, 2006; de Ruiter et al., 2012). In short, gestures alone can serve as turns; they can also be combined with speech. We have made use of this usage in young children in order to unpack some of the steps involved in identifying and then planning a verbal answer.

From Answering Questions to Turn-Taking and Interaction

Turn taking is fundamental in human interaction. Even in infancy, babies respond to parental talk and gaze, initially with extensive overlapping with adult speech in their vocalizations (e.g., Ginsburg and Kilbourne, 1988; Van Egeren et al., 2001). As infants get older, they produce fewer overlaps and more turn-like interactions using gaze, babble, and early word forms (e.g., Rutter and Durkin, 1987; D'Odorici et al., 1997; Hilbrink et al., 2014). But the gaps they leave between turns are often too long. Once they start to use words in their own turns, young children become more adept at anticipating who will talk next when watching an ongoing conversation and they look to the next speaker at appropriate turn-switch points (see Casillas and Frank, under review). But being able to track what is happening in a conversation between third parties is just part of managing interaction. Children also have to learn to take turns on time for themselves.

When people ask questions, they expect to hear answers in the next turn, and the median gap between question and answer for *yes/no* questions in English for adults is around 200 ms (Stivers et al., 2009). But at age one and two, children take considerably longer in producing their answers, although they slowly speed up over the next few years (Casillas et al., under review). And although they get faster in answering simple *yes/no* questions, tracking their increase in speed with age is complicated by the fact that they are simultaneously adding new *Wh*-question words (e.g., *who, how*, *when*, etc.) to their repertoire, and learning how to answer each question type. Yet they clearly know, by age two if not earlier, that they are expected to answer as soon as they can. At this point, they begin to produce floor-holders like *um*, or start an utterance and keep repeating the first word (e.g., *That– that– that–*. . .) until they have their full answer ready (Casillas, 2014b).

Turn-taking is a skill fundamental to language use: it is critical for coordinating with others, whether to exchange greetings, answer questions, exchange information, collaborate in all sorts of activities, or co-construct a story. While some form of turn-taking, with attention to gaze, for example, first emerges during infancy, it is only once children produce recognizable words that they begin to participate in conversational exchanges, and begin trying to observe the adult's 'No gap, no overlap' pattern of taking turns.

Indeed, turn-taking is central to all conversational interaction: it relies on gaze and gesture as well as on the child's utterances. Examining different aspects of turn-taking and, in particular, how children answer different question-types, allows us to take a closer look at how children make use of what they know about both language and interaction so far (Arnon et al., 2014; Grüter and Paradis, 2014). They not only learn language in interaction, but, in interacting, display what they already know and how readily they can process it for production.

Acknowledgments

The preparation of this paper was supported in part by a Stanford Graduate Fellowship to Kate L. Lindsey. We thank Tristan K. Breeden and Rob F. Voigt for help with the data analysis.

References

Andrén, M. (2010). *Children's Gestures from 18 to 30 Months. Centre for Languages and Literature*, Vol. 50. Lund: Lunds Universitet.

Arnon, I., Casillas, M., Kurumada, C., and Estigarribia, B. (eds). (2014). *Language in Interaction: Studies in Honor of Eve V. Clark*. Amsterdam: John Benjamins.

Bangerter, A., and Oppenheimer, D. (2006). Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture* 6, 85–102. doi: 10.1075/gest.6.1.05ban

Bates, E., Camaioni, L., and Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill Palmer Q.* 21, 205–226.

Behne, T., Liszkowski, U., Carpenter, M., and Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *Br. J. Dev. Psychol.* 30, 359–375. doi: 10.1111/j.2044-835X.2011.02043.x

Bruner, J. S. (1975). The ontogenesis of speech acts. *J. Child Lang.* 2, 1–20. doi: 10.1017/S0305000900000866

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63, [Serial No. 255]. doi: 10.2307/1166214

Carter, A. L. (1978). "From sensori-motor vocalizations to words: a case study of the evolution of attention-directing communication on the second year," in *Action, Gesture and Symbol: The Emergence of Language*, ed. A. Lock (London: Academic Press), 309–349.

Caselli, M. C. (1990). "Communicative gestures and first words," in *From Gesture to Language in Hearing and Deaf Children*, eds V. Volterra and C. J. Erting (Berlin: Springer), 56–67. doi: 10.1007/978-3-642-74859-2_6

Casillas, M. (2014a). "Taking the floor on time: delay and deferral in children's turn taking," in *Language in Interaction: Studies in Honor of Eve V. Clark*, eds I. Arnon, M. Casillas, C. Kurumada, and B. Estigarribia (Amsterdam: Benjamins), 101–114.

Casillas, M. (2014b). "Turn-taking," in *Pragmatic Development in First Language Acquisition*, ed. D. Matthews (Amsterdam: Benjamins), 53–70.

Clark, E. V., and Hecht, B. F. (1983). Comprehension, production, and language acquisition. *Ann. Rev. Psychol.* 34, 325–349. doi: 10.1146/annurev.ps.34.020183.001545

Clark, H. H. (2012). "Wordless questions, wordless answers," in *Questions: Formal, Functional and Interactional Perspectives*, ed. J. P. de Ruiter (Cambridge: Cambridge University Press), 81–100. doi: 10.1017/CBO9781139045414.006

Dapretto, M., and Bjork, E. L. (2000). The development of word retrieval abilities in the second year and its relation to early vocabulary growth. *Child Dev.* 71, 635–648. doi: 10.1111/1467-8624.00172

Demuth, K., Culbertson, J., and Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Lang. Speech* 49(Pt 2), 137–174. doi: 10.1177/00238309060490020201

de Ruiter, J. P., Bangerter, A., and Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: investigating the trade-off hypothesis. *Top. Cogn. Sci.* 4, 232–248. doi: 10.1111/j.1756-8765.2012.01183.x

D'Odorici, L., Cassibba, R., and Salerni, N. (1997). Temporal relationships between gaze and vocal behavior in prelinguistic and linguistic communication. *J. Psycholinguist. Res.* 26, 539–556. doi: 10.1023/A:1025027830861

Dromi, E. (1987). *Early Lexical Development*. Cambridge: Cambridge University Press.

Ervin-Tripp, S. (1970). "Discourse agreement: how children answer questions," in *Cognition and the Development of Language*, ed. J. R. Hayes (New York, NY: Wiley), 79–107.

Estigarribia, B., and Clark, E. V. (2007). Getting and maintaining attention in talk to young children. *J. Child Lang.* 34, 349–371. doi: 10.1017/S0305000907008161

Fernald, A., and Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Dev.* 83, 203–222. doi: 10.1111/j.1467-8624.2011.01692.x

Fernald, A., Marchman, V. A., and Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Dev. Sci.* 16, 234–248. doi: 10.1111/desc.12019

Fernald, A., Perfors, A., and Marchman, V. A. (2006). Picking up speed in understanding: speech processing efficiency and vocabulary growth across the 2nd year. *Dev. Psychol.* 42, 98–116. doi: 10.1037/0012-1649.42.1.98

Ginsburg, G. P., and Kilbourne, B. K. (1988). Emergence of vocal alternation in mother-infant interchanges. *J. Child Lang.* 15, 2221–2235. doi: 10.1017/S0305000900012344

Grüter, T., and Paradis, J. (eds). (2014). *Input and Experience in Bilingual Development*. Amsterdam: John Benjamins.

Hilbrink, E., Gattis, M., and Levinson, S. C. (2014). Turn-taking and its timing in infancy: a longitudinal study. *Talk presented at the 4th International Conference on Conversation Analysis* (*ICCA14*). Los Angeles, CA.

Kelly, B. F. (2011). "A new look at redundancy in children's gesture and word combinations," in *Experience, Variation, and Generalization: Learning a First Language*, eds I. Arnon and E. V. Clark (Amsterdam: John Benjamins), 73–90. doi: 10.1075/tilar.7.05kel

Kelly, B. F. (2014). "Temporal synchrony in early multi-modal communication," in *Language in Interaction: Studies in honor of Eve V. Clark*, eds I. Arnon, M. Casillas, C. Kurumada, and B. Estigarribia (Amsterdam: John Benjamins), 117–138.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT.

Levinson, S. C. H. P. (2000). "Grice on location on Rossel Island," in *Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: UCB Department of Linguistics, 210–224.

Levinson, S. C. (2006). "On the human interaction engine," in *Roots of Human Sociality*: *Culture, Cognition and Interaction*, eds N. J. Enfield and S. C. Levinson (Oxford: Berg), 39–69.

Liszkowski, U. (2006). "Infant pointing at 12 months: communicative goals, motives, and social-cognitive abilities," in *Roots of Human Sociality: Culture, Cognition, and Interaction*, eds N. J. Enfield and S. C. Levinson (Oxford: Berg), 153–178.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*, 3rd Edn, Vol. 2. *The Database*. Mahwah, NJ: Lawrence Erlbaum.

McAllister Byun, T., Inkelas, S., and Rose, Y. (2016). The a-map model: articulatory reliability in child-specific phonology. *Language* (in press).

Muñetón Ayala, A. M., and Rodrigo López, M. J. (2011). Functions of the pointing gesture in mothers and their 12- to 36-month-old children during everyday activities. *Spanish J. Psychol.* 14, 619–629. doi: 10.5209/rev_SJOP.2011.v14.n2.10

Olson, J., and Masur, E. F. (2011). Infants' gestures influence mothers' provision of object, action and internal state labels. *J. Child Lang.* 38, 1028–1054. doi: 10.1017/S0305000910000565

Rutter, D. R., and Durkin, K. (1987). Turn-taking in mother-infant interaction: an examination of vocalizations and gaze. *Dev. Psychol.* 23, 54–61. doi: 10.1037/0012-1649.23.1.54

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *J. Child Lang.* 38, 1–34. doi: 10.1017/S0305000910000425

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *J. Mem. Lang.* 60, 252–269. doi: 10.1016/j.jml.2008.11.003

Swingley, D., and Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition* 76, 147–166. doi: 10.1016/S0010-0277(00)00081-0

Tyack, D. L., and Ingram, D. (1977). Children's production and conprehension of questions. *J. Child Lang.* 4, 211–224. doi: 10.1017/S0305000900001616

Van Egeren, L. A., Barratt, M. S., and Roach, M. A. (2001). Mother-infant responsiveness: timing, mutual regulation, and interactional context. *Dev. Psychol.* 37, 684–697. doi: 10.1037/0012-1649.37.5.684

Werner, H., and Kaplan, B. (1963). *Symbol Formation: An Organismic-Developmental Approach to Language and the Expression of Thought*. New York, NY: Wiley.

Wundt, W. (1973). *The Language of Gestures*. The Hague: Mouton. doi: 10.1515/9783110808285

Zangl, R., Klarman, L., Thal, D., Fernald, A., and Bates, E. (2005). Dynamics of word comprehension in infancy: developments in timing, accuracy, and resistance to acoustic degradation. *J. Cogn. Dev.* 6, 179–208. doi: 10.1207/s15327647jcd0602_2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 10 February 2015
doi: 10.3389/fpsyg.2015.00108

The use of intonation for turn anticipation in observed conversations without visual signals as source of information

***Anne Keitel*[1,2]* *and Moritz M. Daum*[2,3]**

[1] *Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK*
[2] *Research Group 'Infant Cognition and Action', Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[3] *Department of Psychology, University of Zurich, Zurich, Switzerland*

Edited by:
Judith Holler, Max Planck Institute Psycholinguistics, Netherlands

Reviewed by:
Marisa Casillas, Max Planck Institute for Psycholinguistics, Netherlands
Christine Fawcett, Uppsala University, Sweden

***Correspondence:**
Anne Keitel, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK
e-mail: anne.keitel@glasgow.ac.uk

The anticipation of a speaker's next turn is a key element of successful conversation. This can be achieved using a multitude of cues. In natural conversation, the most important cue for adults to anticipate the end of a turn (and therefore the beginning of the next turn) is the semantic and syntactic content. In addition, prosodic cues, such as intonation, or visual signals that occur before a speaker starts speaking (e.g., opening the mouth) help to identify the beginning and the end of a speaker's turn. Early in life, prosodic cues seem to be more important than in adulthood. For example, it was previously shown that 3-year-old children anticipated more turns in observed conversations when intonation was available compared with when not, and this beneficial effect was present neither in younger children nor in adults (Keitel et al., 2013). In the present study, we investigated this effect in greater detail. Videos of conversations between puppets with either normal or flattened intonation were presented to children (1-year-olds and 3-year-olds) and adults. The use of puppets allowed the control of visual signals: the verbal signals (speech) started exactly at the same time as the visual signals (mouth opening). With respect to the children, our findings replicate the results of the previous study: 3-year-olds anticipated more turns with normal intonation than with flattened intonation, whereas 1-year-olds did not show this effect. In contrast to our previous findings, the adults showed the same intonation effect as the 3-year-olds. This suggests that adults' cue use varies depending on the characteristics of a conversation. Our results further support the notion that the cues used to anticipate conversational turns differ in development.

Keywords: turn-taking, conversations, intonation, visual cues, interaction, infants, adults, eye tracking

INTRODUCTION

Smooth and successful everyday social interactions are to a large extent based on the individual's ability to anticipate what an interaction partner is going to do next. The anticipation of the next step is crucial for successful non-verbal interactions, for example, in sport or musical performances, but also for successful verbal interactions, that is, conversations. During a conversation, a variety of cues such as content or prosody mark the end of one speaker's turn and the beginning of the next speaker's turn. The present study was designed to explore the impact of two non-symbolic cues on the identification of a speaker's turn in greater detail, with a particular focus on the development of this cue use: the prosodic cue intonation and visual signals that are unrelated to linguistic content and form.

When adults engage in a conversation, they identify turn transitions without great effort. They can use a variety of cues to do so: (1) the semantic content of a turn, or lexico-syntactic information, indicates that a response is required (*content cues*), (2) the spoken content is modulated by prosodic cues, such as intonation, to indicate a turn-boundary (*prosodic cues*), and (3) visual information peripherally accompanies speech, such as opening the mouth or gestures (*visual cues*). The use of lexico-syntactic content is highly related to language comprehension and seems to be among the most important factors for detecting the end of a turn. For example, de Ruiter et al. (2006) presented audio recordings of isolated turns from natural telephone conversations to adult participants, and asked them to indicate the anticipated end of the speaker's turn. Participants reliably indicated an expected end of a turn well before the turn was actually finished. A follow-up study suggested that participants anticipated the upcoming lexical content and used this information to estimate the end of a turn (Magyari and de Ruiter, 2012). Accordingly, adults not only accurately detect a turn boundary, they usually anticipate it if they can rely on the spoken content.

Spoken language, however, not only includes semantic and syntactic cues but is also accompanied by rich non-symbolic information. Other linguistic information includes acoustically marked prosodic boundary cues (Gerken, 1996) that involve

intonation, syllable length, and pauses. The availability of these cues helps to segment linguistic units even when information about content is unavailable (Schaffer, 1983; de Ruiter et al., 2006; Casillas and Frank, 2012). For example, when utterances are made unintelligible, with only prosodic cues (notably intonation) still intact, adults can identify the end and beginning of turns at above chance level (Schaffer, 1983; de Ruiter et al., 2006). In the reverse case, when utterances lack prosodic information, the ability to recognize the end of a turn is similar to when the full repertoire of information is available (de Ruiter et al., 2006). Thus, adults' ability to detect an end of turn is not influenced by the availability of intonation (de Ruiter et al., 2006). These findings have led to the notion that adults primarily use prosody to predict the end of a turn if semantic and syntactic information is lacking (Grosjean and Hirt, 1996).

Visual information also contributes to turn taking in natural adult conversation. This includes visual cues such as mouth opening before speech onset, language-accompanying body movements and gesture, as well as gaze (Thórisson, 2002). For example, analyses of visual signals in conversations from a previous study by Keitel et al. (2013) yielded results indicating that the speakers opened their mouths on average 494.43 ms (SD = 228.48 ms) before a verbal speech signal was audible. Thus, it is well possible that visual cues support the detection of a turn. Taken together, adults make use of a variety of cues to detect turns during conversations, of which lexico-syntactic content seems to play a major role, and prosodic and visual information serve a supportive function.

Early in life, when language skills are far from adult level, children lack a substantial amount of the linguistic repertoire required for identifying the end of a turn. Studies on language development suggest that infants' word comprehension starts at around the age of 8 months and rapidly increases over the next few months so that their productive vocabulary has reached approximately 600 words by the age of 30 months (Fenson et al., 1994). During this time, the child's vocabulary not only increases substantially, but their sentences also become more complex (Clark, 2009). This development results in a rather sophisticated understanding and application of syntactic schemes at around age 3.5–4 years (Tomasello, 2000).

Regarding the anticipation of turns, eye tracking studies have shown that, with increasing language skills, children increasingly anticipate turns in observed conversations. One of the first studies that addressed this topic analyzed the gaze pattern of 1- and 3-year-old children during the observation of an everyday conversation between two speakers (von Hofsten et al., 2009). The authors of the study analyzed whether the observing children shifted their gaze to the next speaker before they started speaking. These results showed that the anticipation of turns increased significantly with age. Findings from similar eye tracking studies have shown that 1- to 7-year-old children anticipate speaker transitions in observed conversations effectively (Casillas and Frank, 2012, 2013). Furthermore, and most importantly for the present study, recent findings from Keitel et al. (2013) extended these results with a reliability analysis of participants' gaze shifts. This analysis revealed that the anticipation of turns was reliable only in 3-year-olds and adults. Younger children shifted their gaze between speakers mostly independently of the turn-taking. These findings suggest that children need a sophisticated level of language understanding to anticipate conversations in a similar manner to adults, which is acquired around the age of 3 years.

Early in life, prosodic cues serve particularly important functions in children's language development. For example, prosody helps 6-month-old infants to segment linguistic units, such as clauses (Nazzi et al., 2000) and phrases (Soderstrom et al., 2003). Furthermore, the prosodic cue intonation can already be extracted from speech by newborns (Nazzi et al., 1998; Sambeth et al., 2008). For the anticipation of turns in observed conversations, the role of intonation was investigated in the above-mentioned study by Keitel et al. (2013). In this study, children between 6 months and 3 years of age, as well as adults, saw videos of two dyadic conversations. The auditory signal of the conversations, in particular, intonation, was modulated. In one condition intonation was kept normal; in a second condition intonation was synthetically flattened. Interestingly, only the 3-year-old children benefitted from the unmodified and available intonation. Neither the adults nor the younger children showed differences in their gaze behavior between the two conditions. The lack of an intonation effect in younger children could be due to the fact that they often shifted their gaze between speakers unrelated to turn transitions. Nevertheless, intonation seems to be important for 3-year-olds to anticipate the course of observed conversations. However, a different study that investigated the role of lexical and prosodic information on 1- to 7-year-old children's turn anticipation did not find this effect in 3-year-olds (Casillas and Frank, 2013). Casillas and Frank (2012, 2013) additionally differentiated between gaze shifts following questions or non-questions and found an advantage for questions in older children (beginning around 3–4 years). They conclude that children's turn anticipation relies on both lexical and prosodic information. Thus, both studies, Keitel et al. (2013) and Casillas and Frank (2013), suggest that children use both lexical and prosodic cues to identify the end of turns. In contrast, studies with adults have concluded that they predominantly rely on the information provided by the lexical content (e.g., de Ruiter et al., 2006).

In the present study, we investigated the effect of the prosodic cue intonation in children and adults while controlling for visual cues (see also Casillas and Frank, 2013). During natural conversations between two human interaction partners, diverse visual cues can indicate an upcoming turn transition. For example, the mouth is usually opened before the actual stream of speech starts, and small gestures can also indicate a desire to respond. To avoid these visual cues entirely, we presented videos of conversations between two puppets. The puppets did not move their bodies, and the onset of the speech signal perfectly corresponded with the onset of the visual signal (mouth opening). The same was true for the offset of the acoustic and visual speech signals. Again, as in the study by Keitel et al. (2013), intonation was either normal or flattened. In addition to an adult control sample, we tested children aged 1 year, just starting their first words, and children aged 3 years, fluently using multi-phrasal utterances (von Hofsten et al., 2009; Keitel et al., 2013). The aim of the current study was to corroborate previous findings by Keitel et al. (2013) and von Hofsten et al. (2009) while exploring the impact of missing

visual cues. Based on previous findings, we expected an increase of turn anticipations with age and a beneficial effect of available intonation at age 3.

MATERIALS AND METHODS

PARTICIPANTS

Seventy-two participants, 24 in each of the three age groups, were included in the final analyses: 1-year-old children [15 female, 9 male; *M*(age) = 12 months 4 days, range = 11 months 16 days to 12 months 15 days], 3-year-old children [13 female, 11 male; *M*(age) = 36 months 0 days, range = 35 months 17 days to 36 months 15 days], and adults [11 female, 13 male; *M*(age) = 23.5 years, range = 20–30 years]. Ten additional 1-year-olds and three additional 3-year-olds were tested but excluded from analysis because they were inattentive towards the conversations and did not yield enough valid trials (see Data Analysis for valid trials criteria). One additional adult participant was tested but was excluded from data analysis due to a technical error. Contact information of children was obtained from public birth records from the city of Leipzig, Germany. Children received a toy and adults received monetary compensation for their participation. The study was approved by a local ethics committee at the University of Leipzig and conducted in accordance with the Declaration of Helsinki.

APPARATUS AND STIMULI

Two different conversations between animal puppets were presented (see **Figure 1A**). The topics were recreational activities (conversation A) and birthday plans (conversation B). Each consisted of 28 turns that were analyzed (i.e., 27 speaker switches, or trials). The puppets first greeted the participants directly (these turns were not included in the analyses). The number of questions and declarative sentences were identical for each actor and conversation. The average length of speech and gaps differed only by approximately 295 (i.e., 11.9%) and 28 ms (3.1%), respectively, between conditions (see **Table 1** for details of the conversations). The average duration of gaps in our study is longer than in a typical adult conversation (915 ms in our study vs. approximately 400–500 ms in other studies; see Heldner and Edlund, 2010). We decided to use distinct gaps to give even the younger children enough time to process them.

To create the conversations, audio tracks were recorded first. Two female actors held the scripted conversations using normal, adult-directed intonation and spoke the lines into a microphone. Subsequently, the conversations were acted out using hand puppets that could open and close their mouths independently from the rest of the body. The audio recordings were played back during the video recording of the puppets so that the actors could move the puppet's mouth synchronously with the speech signal. Audio and video tracks were then combined using video editing software (Edius). The tracks were arranged in a way that the movement of the mouth started and finished at exactly the same time as the speech. For both conversations, a second version was created with flattened intonation (using the same procedure as in Keitel et al., 2013). For the flattened intonation conversations, the variations of the fundamental frequency (F0) were removed and averaged to the mean frequency of the conversations using the software Praat (Boersma and Weenink, 2014). Via this procedure, the pitch contour of the conversations was extracted and segmented into pitch points at a rate of 100 Hz. The pitch points were removed, and a new pitch contour was created with the average frequency of the respective conversation using pitch synchronous overlap and add (PSOLA) resynthesis. This resulted in clearly less intonated, monotone speech (see **Figure 1B** for pitch properties). Thus, both conversations A and B were available in the normal condition and in the flattened condition.

Table 1 | Details of the two conversations. Number of analyzed speaker transitions, mean number of words per turn, total duration of conversations, mean duration of speech (i.e., mean duration of individual turn utterances), and mean duration of inter-turn gaps.

	Number of analyzed speaker switches	Ø Words/ Turn	Duration (in seconds)		
			Total	Ø Speech	Ø Gaps
Conversation A	27	7.9	88.24	2.34	0.93
Conversation B	27	9.3	95.44	2.63	0.90

The videos were presented on a 17-inch monitor (resolution: 800 × 600 pixels). They subtended a visual angle of 27.5° × 15.2°. The participants sat at a distance of approximately 60 cm from the monitor. Gaze was measured using a remote corneal reflection eye tracker (Tobii 1750, Stockholm, Sweden, with infant add-on; sampling rate: 50 Hz; precision: 1°; accuracy: 0.5°; software: ClearView 2.7.1). The stereo audio signal was played back via two speakers placed either side of the monitor.

PROCEDURE

The experimenter explained the procedure to the children and their parents and to the adult participants; otherwise, participants were naïve to the purpose of the study. Written informed consent was obtained from the adult participants and from the children's parents. After the child had become accustomed to the experimenter, the experiment started. Each participant was tested individually; one parent was present during testing. A 9-point infant calibration was used for all participants (this took approximately 30 s to 1 min) before the conversation videos were presented. Each participant watched both conversations, A and B, one with normal intonation and one with flattened intonation (this took approximately 3.5 min). The order of the conversations and the intonation were counterbalanced across participants. Before each video presentation, an attention-grabbing video (including interesting toys that moved and made sounds) was shown to the participants to focus their attention on the monitor.

DATA ANALYSIS

Eye movement data were analyzed using the software Matlab R2013b (The MathWorks). To detect gaze shifts towards speakers, areas of interest (AOIs) were defined that covered each puppet's mouth and eyes (see white boxes in **Figure 1A**). For the two speakers in conversation A, AOIs covered an area of 5.8° × 6.1°, and 6.0° × 7.2° visual angle, respectively. In conversation B, AOIs

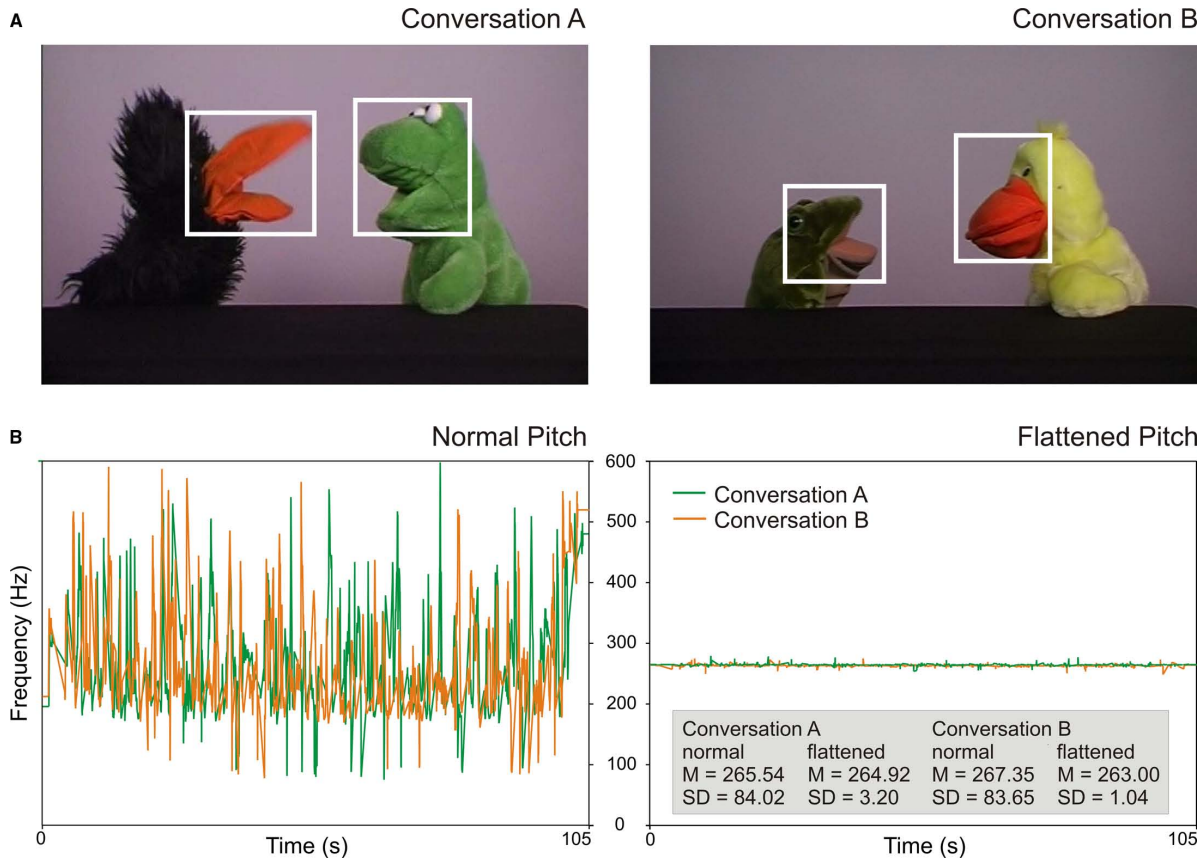


FIGURE 1 | (A) Screenshots of the two presented conversations. The white boxes illustrate areas of interest for the analysis of gaze shifts towards speakers. **(B)** Pitch properties of the two conversations. Lines display the F0 contours of the normal conversations (left) and of the manipulated, flattened conversations (right). Mean pitch and standard deviations (in Hz) are indicated in the gray box.

covered an area of 5.1° × 5.5°, and 4.9° × 6.5° visual angle, respectively.

Three measures were calculated to characterize participants' gaze behavior. First, the exact time that gaze arrived at the next speaker relative to the beginning of their turn was calculated (*gaze latency*). A gaze shift towards the next speaker was considered to be anticipatory when it had arrived at the speaker before they had begun to speak, and reactive when it arrived after they had started speaking. Positive values indicate an anticipatory gaze shift; negative values indicate a reactive gaze shift. Second, we analyzed the location and duration of fixations provided by the data acquisition software (*fixation duration*; Keitel et al., 2014). Fixation duration can indicate distraction. For example, shorter fixation durations in the flattened condition could suggest that participants were distracted by the unusual intonation. Third, we calculated the occurrences of anticipatory and random gaze shifts (*occurrence rate*; Keitel et al., 2013). Occurrence rates are calculated as the number of gaze shifts during specific time intervals. For anticipatory gaze shifts, these time intervals were inter-turn gaps, including the final 500 ms prior to the end of a turn. The direction of anticipatory gaze shifts was always towards the (upcoming) speaker. For random gaze shifts, time intervals included speech (minus the 500-ms gap at the end of a turn). The direction of random gaze shifts was always away from the speaker (see Keitel et al., 2013 for detailed illustration of time intervals and direction of gaze shifts). Because occurrence rates are calculated as the number of occurrences per time interval, they can be interpreted as a probability to make a gaze shift. If the probability to make anticipatory gaze shifts is statistically higher than the probability to make random gaze shifts, turn anticipation is considered reliable.

Gaze shifts towards a speaker were only regarded valid if they were immediately preceded by a 100-ms fixation on the other speaker. This limitation was included to ensure that a gaze shift was related to the conversation and not random (Keitel et al., 2013, 2014, see also Melzer et al., 2012).

To be included in the analyses, a participant had to show at least 10 turn-taking-related gaze shifts (i.e., either anticipatory or reactive) in each condition, out of the 27 possible trials. In the group of 1-year-olds, children showed an average of 18.42 (SD = 4.72; normal condition), and 19.29 (SD = 4.20; flattened condition) valid trials, respectively. The 3-year-old children showed an average of 20.88 (SD = 4.37; normal condition), and 21.46 (SD = 3.90; flattened condition) valid trials, respectively.

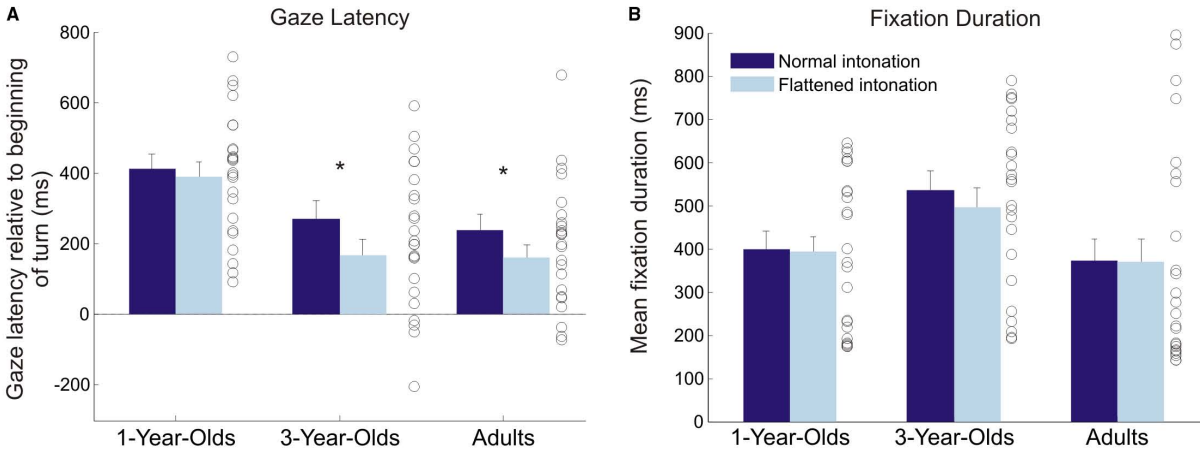


FIGURE 2 | Average gaze latency. (A) and fixation duration **(B)** in the normal and flattened condition for all age groups. Error bars show standard error of the mean. Circles illustrate individual values for each participant averaged over both conditions. Asterisks indicate significant difference between conditions. Time point zero on the y-axis in the left plot refers to the beginning of a turn in the video. Positive values indicate that gaze arrived at the speaker before they started speaking; negative values indicate that gaze arrived after they had started speaking.

Adults showed an average of 24.75 (SD = 2.38; normal condition), and 24.71 (SD = 2.35; flattened condition) valid trials, respectively. Paired *t*-tests with number of valid trials between conditions did not suggest a difference for any age group (all *p*s > 0.38, two-sided). The results presented below are the same even with a simpler inclusion criterion—gazing at the screen at least 50% of the time in total—so it is unlikely that our inclusion criterion of 10 turn-taking-related gaze shifts introduced bias to our findings.

RESULTS

GAZE LATENCY

Initial analyses did not suggest any main effect or interaction effects of video order on gaze latency (all *p*s > 0.31), and data were collapsed over this factor. In all age groups and conditions, participants showed positive mean gaze latencies, which means, on average, they anticipated turns (*t*-tests against zero; 1-year-olds: normal condition: $t(23) = 9.40$, $p < 0.001$; flattened condition: $t(23) = 9.37$, $p < 0.001$; 3-year-olds: normal condition: $t(23) = 5.53$, $p < 0.001$; flattened condition: $t(23) = 3.98$, $p = 0.001$; adults: normal condition: $t(23) = 5.32$, $p < 0.001$; flattened condition: $t(23) = 4.79$, $p < 0.001$; see **Figure 2A**).

A 3 × 2 (age [1 year, 3 years, adults]) × condition [normal, flattened]) analysis of variance (ANOVA) with gaze latency yielded significant main effects of age, $F(2,69) = 9.52$, $p < 0.001$, $\eta_G^2 = 0.17$, and condition, $F(1,69) = 7.33$, $p = 0.009$, $\eta_G^2 = 0.02$, and no significant interaction, $F < 1$ (generalized eta-squared values are given to facilitate comparability with other studies, see Olejnik and Algina, 2000; Bakeman, 2005). Bonferroni-corrected *post hoc* *t*-tests showed that the 1-year-olds shifted their gaze earlier than the 3-year-olds, $p = 0.001$, and earlier than adults, $p = 0.001$. There was no significant difference between the gaze latencies of the 3-year-olds and the adults, $p = 1$. Following up the effect of condition, paired *t*-tests showed that 3-year-olds, $t(23) = 2.30$, $p = 0.03$, $d = 0.47$, as well as adults, $t(23) = 2.17$, $p = 0.04$, $d = 0.44$, displayed earlier gaze shifts in the normal than in the flattened condition, whereas 1-year-olds did not show this effect, $t(23) = 0.47$, $p = 0.64$, $d = 0.10$.

DISTRIBUTION AND DURATION OF FIXATIONS

Figure 3 illustrates the fixation distribution in conversation A for both conditions (see **Figure 2B** for means of both conversations). The example illustrates similarly focused fixations on the puppets' faces in both conditions. A 3 × 2 (age × condition) ANOVA with fixation duration yielded a significant main effect of age, $F(2,69) = 3.26$, $p = 0.045$, $\eta_G^2 = 0.08$, and no significant main effect of, or interaction effect with, intonation (both $F < 1$; see **Figure 2B**). Bonferroni-corrected *post hoc* *t*-tests showed that 3-year-olds had marginally longer fixation durations than adults, $p = .06$.

OCCURRENCE RATE OF ANTICIPATORY AND RANDOM GAZE SHIFTS

The occurrence probabilities, or occurrence rates (see **Figure 4**), to make either anticipatory or random gaze shifts were entered into a 3 × 2 [age (1 year, 3 years, adults) × occurrences (anticipatory, random)] ANOVA. Results yielded main effects of age, $F(2,69) = 31.17$, $p < 0.001$, $\eta_G^2 = 0.38$, and occurrences, $F(1,69) = 116.07$, $p < 0.001$, $\eta_G^2 = 0.20$, and a significant interaction between both, $F(2,69) = 15.06$, $p < 0.001$, $\eta_G^2 = 0.06$. Bonferroni-corrected *post hoc* *t*-tests showed significant differences between 1-year-olds and both older age groups, both $p < 0.001$, but not between 3-year-olds and adults, $p = 0.51$. All age groups indicated larger occurrence rates for anticipatory gaze shifts than for random gaze shifts, but this difference was only significant in 3-year-olds, $t(23) = 6.54$, $p < 0.001$, $d = 1.35$, and adults, $t(23) = 11.40$, $p < 0.001$, $d = 2.33$. The 1-year-olds showed only marginally higher occurrence rates for anticipatory gaze shifts, $t(23) = 1.92$, $p = 0.07$, $d = 0.39$. Moreover, the rates for random gaze shifts decreased with age (all comparisons,

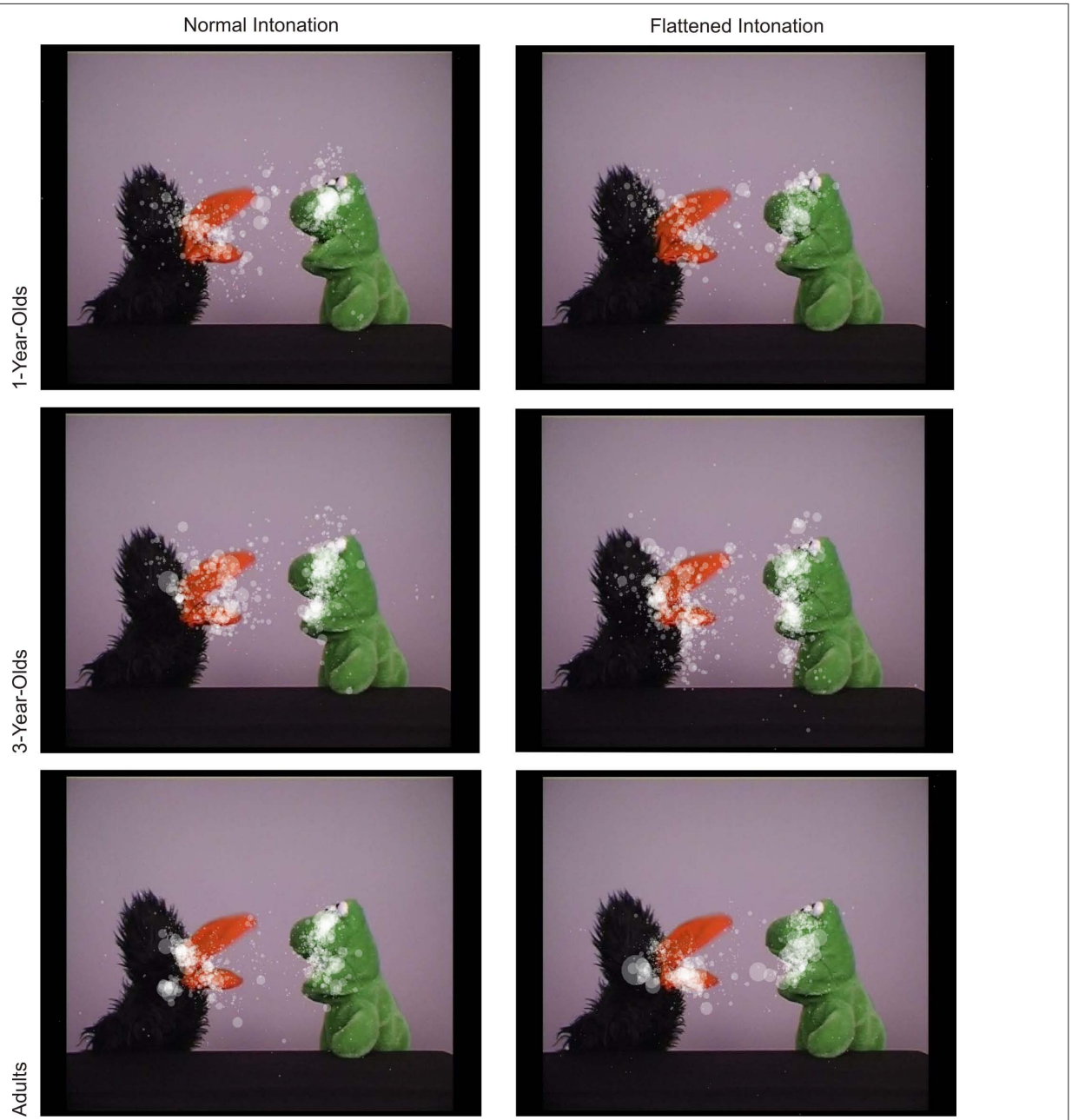


FIGURE 3 | Distribution of fixations in the normal and flattened condition (conversation A only) for all age groups. Each transparent dot displays a fixation; its size indicates the fixation duration.

$p < 0.005$); the rates for anticipatory gaze shifts differed only between 1-year-olds and older age groups, both $p < 0.003$, but not between 3-year-olds and adults, $p = 0.92$.

DISCUSSION

CHILDREN'S AND ADULTS' CUE USE FOR TURN ANTICIPATION

In the present study we investigated the effect of intonation on turn anticipation during the observation of a dyadic conversation while controlling the availability of visual cues. The main finding was that 3-year-olds and adults anticipated more turns with normal intonation than with flattened intonation, whereas 1-year-olds did not show this effect. The effect of the children's data replicates previous findings that 3-year-olds but not 1-year-olds benefit from the additionally available intonation (Keitel et al., 2013, but see Casillas and Frank, 2013). This was, as in the earlier study, not caused by general differences in the allocation of attention towards the conversation because fixation durations did not differ between conditions. In line with previous findings, we argue that at age 3, children have, on the one hand, learned to use prosodic boundary cues to indicate higher level

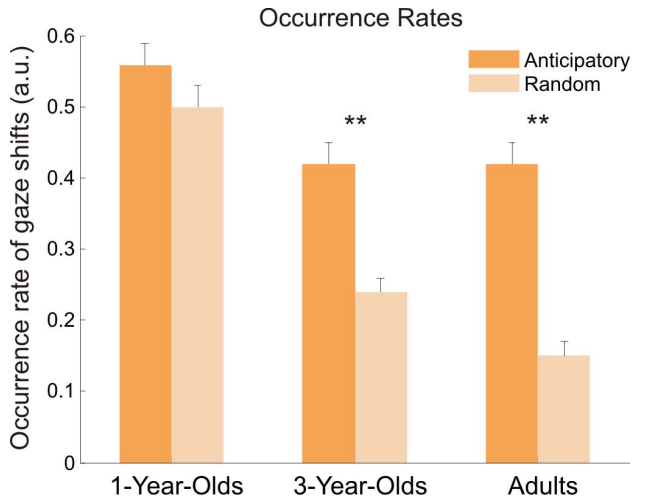


FIGURE 4 | Occurrence rates for anticipatory and random gaze shifts for all age groups. Occurrence rates represent a probability to make a gaze shift to the other speaker. Error bars show standard error of the mean. Asterisks indicate significant difference between the rates for anticipatory and random gaze shifts.

linguistic aspects (Männel and Friederici, 2011) but, on the other hand, their overall language skills are not yet as sophisticated as in adults (e.g., Clark, 2009). Accordingly, at age 3, the information provided by intonation effectively supports the perception of conversations and helps to anticipate a speaker's next turn.

The availability of the prosodic cue intonation yielded an increase of turn anticipations also in the adult participants. Usually, during the observation of normal conversations, adults heavily rely on lexico-syntactic cues (de Ruiter et al., 2006; Magyari and de Ruiter, 2012) and they only make use of prosodic cues when lexico-syntactic information is lacking (Grosjean and Hirt, 1996). The present findings, however, indicate that adults' cue use is even more flexible: prosodic cues might be beneficial not only with lacking lexico-syntactic cues but also with lacking visual cues. This also gives rise to the assumption that adults naturally use visual cues in conversations to detect a speaker's intention to respond. However, the lack of visual cues did not have a drastic effect on participants' turn anticipation, as, on average, they anticipated an upcoming turn in both conditions. This supports at least the assumption that visual cues are not mandatory for turn anticipation.

EXCEPTIONAL TURN ANTICIPATION IN 1-YEAR-OLDS

A second, somewhat unexpected, finding was that 1-year-olds anticipated more turns than older participants, independent of condition. The distribution of individual means in **Figure 2** illustrates that these results were not caused by outliers, but that 1-year-olds were consistently good at anticipating the course of the conversations. However, the analysis of occurrence rates for anticipatory and random gaze shifts helps to interpret this finding: 1-year-olds generally showed a higher probability for making gaze shifts than 3-year-olds and adults. Importantly, there was no significant difference between the probabilities for making random and anticipatory gaze shifts, which suggests that 1-year-olds' turn anticipation was not yet reliable. Furthermore, the probability of making random gaze shifts decreased significantly with age. This finding suggests that younger children gazed back and forth between the speakers much more often than older children and adults. The histograms of gaze latencies (**Figure 5**) illustrate that older participants' gaze shifts to the next speaker center around the turn onset, whereas 1-year-olds' gaze shifts to the next speaker are more evenly distributed over the whole time interval. An appropriate interpretation of these findings is that 1-year-olds shifted their gaze back and forth between speakers for the whole duration of the conversation, and this resulted in high probabilities of random *and* anticipatory gaze shifts. These constant gaze shifts could, on the one hand, be due to shorter attention spans in young children compared to older children and adults. On the other hand, puppets could have been particularly interesting for the 1-year-olds, resulting in keen visual exploration.

ECOLOGICAL VALIDITY OF PUPPET CONVERSATIONS

The use of puppets as conversation partners in the present study raises the question of ecological validity and the tendency to generalize the present results to the "real world." For human conversation partners, it is not possible to sit perfectly still or avoid mouth opening prior to speaking, without seeming unnatural or robotic. Therefore, the use of puppets to solve this problem seems justified (see also Casillas and Frank, 2013). Furthermore, studies have shown that young children and adults readily ascribe human qualities to non-human agents, even if they consist of geometrical shapes (Montgomery and Montgomery, 1999; Abell et al., 2000). Puppets should therefore make it easy for observers to immerse themselves in watching the conversations similar to a human conversation. This assertion is supported by the replication of results in 1- and 3-year-olds, compared with the findings of Keitel et al. (2013). However, a little uncertainty might remain that the adult findings are not due to the lack of visual signals but to the use of puppets. But even if this were the case, it would not affect the general conclusion of the present study that available information in a conversation results in differential cue use by adults. To partly resolve this issue, further studies could use puppets as conversation partners and include the typical lag between visual cues and the verbal speech signal.

CONCLUSION

We investigated the effect of intonation on children's and adults' turn anticipation during the observation of dyadic conversations between puppets. When visual cues were lacking, both adults and 3-year-olds benefitted from the availability of intonation. Considering that adults did not show an intonation effect when visual cues were available (Keitel et al., 2013), this suggests that their cue use is rather flexible, depending on available information. Our results demonstrate further developmental differences in the perception of conversations: one-year-olds showed generally more gaze shifts when observing conversations than 3-year-olds and adults, and did not yet anticipate turns reliably. One the

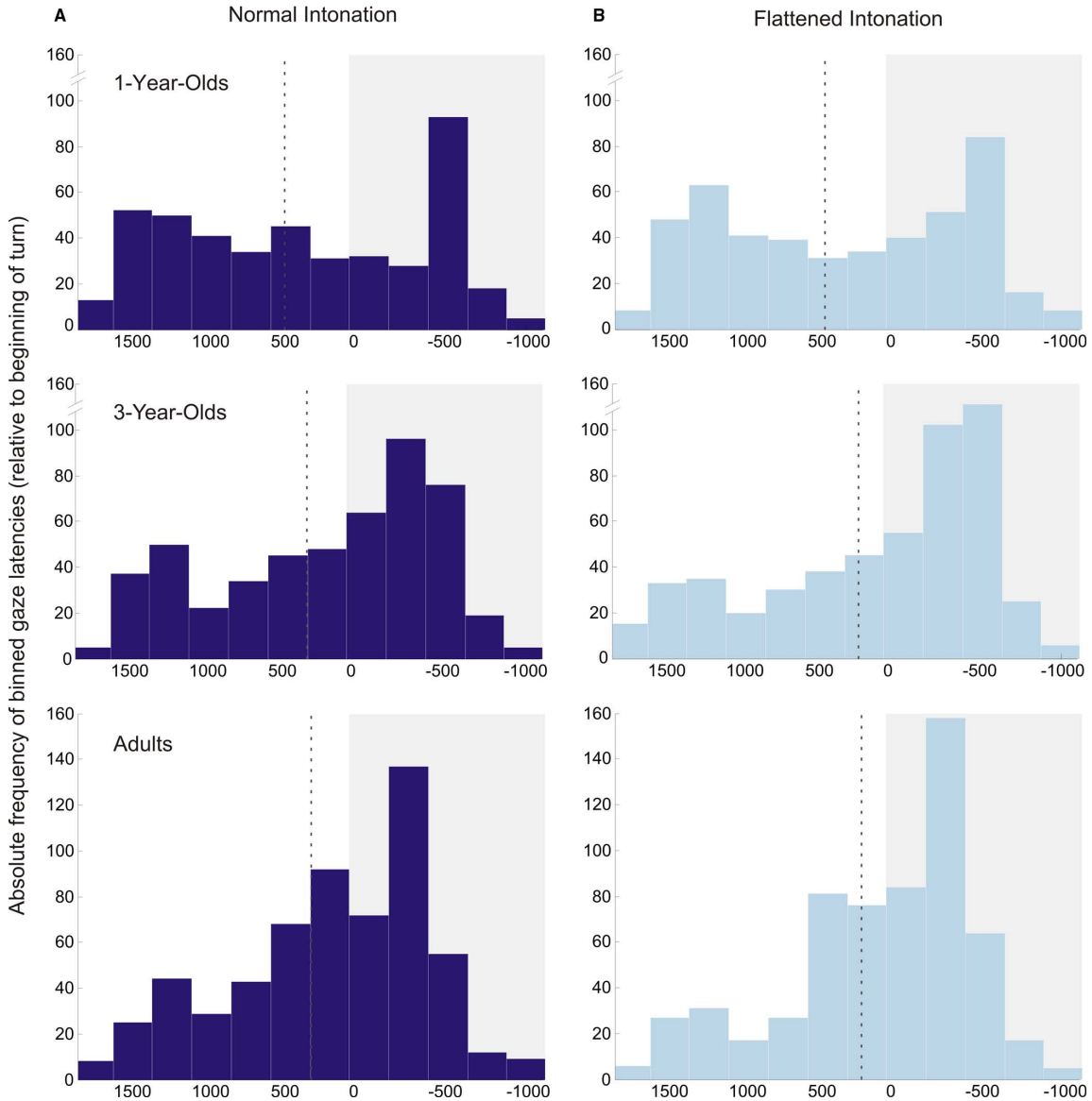


FIGURE 5 | Histograms of anticipatory and reactive gaze shifts in the normal (A) and flattened (B) condition for all age groups. Values during a time interval with a white background illustrate anticipatory gaze shifts; values during a time interval with a gray background illustrate reactive gaze shifts (zero indicates beginning of turn). The dotted line in each plot displays the average gaze latency in this age group/condition. Bin size is 250 ms.

one hand, this makes interpretations of their cue use for turn anticipation difficult. On the other hand, a more fine-grained investigation into 1-year-olds' many "random" gaze shifts might lead to a better understanding of the factors that influence their perception of conversations.

ACKNOWLEDGMENTS

We would like to thank Hagen Lipka for data recovery and Elizabeth Kelly for proofreading. Anne Keitel gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft (Graduate School "Function of Attention in Cognition," GRK 1182) and from the University of Leipzig ("Wiedereinstiegsstipendium").

REFERENCES

Abell, F., Happé, F., and Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cogn. Dev.* 15, 1–16. doi: 10.1016/S0885-2014(00)00014-9

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behav. Res. Methods* 37, 379–384. doi: 10.3758/BF03192707

Boersma, P., and Weenink, D. (2014). *Praat: Doing Phonetics by Computer* [Computer program], Version 5.1.32. Retrieved from http://www.praat.org/

Casillas, M., and Frank, M. C. (2012). "Cues to turn boundary prediction in adults and preschoolers," in *Proceedings of SemDial 2012 (SeineDial): the 16th Workshop on the Semantics and Pragmatics of Dialogue*, eds S. Brown-Schmidt, J. Ginzburg, and S. Larsson (Paris: Université Paris-Diderot), 61–69.

Casillas, M., and Frank, M. C. (2013). "The development of predictive processes in children's discourse understanding," in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Society), 299–304.

Clark, E. V. (2009). *First Language Acquisition*. Cambridge, MA: Cambridge University Press.

de Ruiter, J.-P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, i+iii–v+1–185. doi: 10.2307/1166093

Gerken, L. (1996). Prosody's role in language acquisition and adult parsing. *J. Psycholinguist. Res.* 25, 345–356. doi: 10.1007/BF01708577

Grosjean, F., and Hirt, C. (1996). Using prosody to predict the end of sentences in English and French: normal and brain-damaged subjects. *Lang. Cogn. Process.* 11, 107–134. doi: 10.1080/016909696387231

Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *J. Phon.* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002

Keitel, A., Prinz, W., and Daum, M. M. (2014). Perception of individual and joint action in infants and adults. *PLoS ONE* 9:e107450. doi: 10.1371/journal.pone.0107450

Keitel, A., Prinz, W., Friederici, A. F., von Hofsten, C., and Daum, M. M. (2013). Perception of conversations—the importance of semantics and intonation in children's development. *J. Exp. Child Psychol.* 16, 264–277. doi: 10.1016/j.jecp.2013.06.005

Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

Männel, C., and Friederici, A. D. (2011). Intonational phrase structure processing at different stages of syntax acquisition: ERP studies in 2-, 3-, and 6-year-old children. *Dev. Sci.* 14, 786–798. doi: 10.1111/j.1467-7687.2010.01025.x

Melzer, A., Prinz, W., and Daum, M. M. (2012). Production and observation of contralateral reaching: a close link by 12 months of age. *Infant Behav. Dev.* 35, 570–579. doi: 10.1016/j.infbeh.2012.05.003

Montgomery, D. E., and Montgomery, D. A. (1999). The influence of movement and outcome on young children's attributions of intention. *Br. J. Dev. Psychol.* 17, 245–261.

Nazzi, T., Floccia, C., and Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behav. Dev.* 21, 779–784. doi: 10.1016/S0163-6383(98)90044-3

Nazzi, T., Nelson, D. G., Jusczyk, P. W., and Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: effects of prosodic well-formedness. *Infancy* 1, 123–147. doi: 10.1207/S15327078IN0101_11

Olejnik, S., and Algina, J. (2000). Measures of effect size for comparative studies: applications, interpretations, and limitations. *Contemp. Educ. Psychol.* 25, 241–286. doi: 10.1006/ceps.2000.1040

Sambeth, A., Ruohio, K., Alku, P., Fellman, V., and Huotilainen, M. (2008). Sleeping newborns extract prosody from continuous speech. *Clin. Neurophysiol.* 119, 332–341. doi: 10.1016/j.clinph.2007.09.144

Schaffer, D. B. (1983). The role of intonation as a cue to turn taking in conversation. *J. Phon.* 11, 243–257.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., and Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: evidence from prelinguistic infants. *J. Mem. Lang.* 49, 249–267. doi: 10.1016/S0749-596X(03)00024-X

Thórisson, K. R. (2002). "Natural turn-raking needs no manual: computational theory and model, from perception to action," in *Multimodality in Language and Speech Systems*, eds B. Granström, D. House, and I. Karlsson (Netherlands: Springer), 173–207. Retrieved from http://link.springer.com/chapter/10.1007/978-94-017-2367-1_8

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition* 74, 209–253. doi: 10.1016/S0010-0277(99)00069-4

von Hofsten, C., Uhlig, H., Adell, M., and Kochukhova, O. (2009). How children with autism look at events. *Res. Autism Spectr. Disord.* 3, 556–569. doi: 10.1016/j.rasd.2008.12.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 September 2014; accepted: 21 January 2015; published online: 10 February 2015.

Citation: Keitel A and Daum MM (2015) The use of intonation for turn anticipation in observed conversations without visual signals as source of information. Front. Psychol. 6:108. doi: 10.3389/fpsyg.2015.00108

This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 April 2015
doi: 10.3389/fpsyg.2015.00495

Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions

Imme Lammertink[1]*, *Marisa Casillas*[2], *Titia Benders*[3], *Brechtje Post*[4] and *Paula Fikkert*[1]

[1] Centre for Language Studies, Radboud University, Nijmegen, Netherlands, [2] Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [3] School of Psychology, Newcastle University, Newcastle, NSW, Australia, [4] Department of Theoretical and Applied Linguistics, Cambridge University, Cambridge, UK

Adults achieve successful coordination during conversation by using prosodic and lexicosyntactic cues to predict upcoming changes in speakership. We examined the relative weight of these linguistic cues in the prediction of upcoming turn structure by toddlers learning Dutch (Experiment 1; $N = 21$) and British English (Experiment 2; $N = 20$) and adult control participants (Dutch: $N = 16$; English: $N = 20$). We tracked participants' anticipatory eye movements as they watched videos of dyadic puppet conversation. We controlled the prosodic and lexicosyntactic cues to turn completion for a subset of the utterances in each conversation to create four types of target utterances (*fully incomplete, incomplete syntax, incomplete prosody,* and *fully complete*). All participants (Dutch and English toddlers and adults) used both prosodic and lexicosyntactic cues to anticipate upcoming speaker changes, but weighed lexicosyntactic cues over prosodic ones when the two were pitted against each other. The results suggest that Dutch and English toddlers are already nearly adult-like in their use of prosodic and lexicosyntactic cues in anticipating upcoming turn transitions.

Keywords: turn prediction, prosody, lexicosyntax, child language, eye-tracking

OPEN ACCESS

Edited by:
Nuria Sebastian-Galles, Universitat Pompeu Fabra, Spain

Reviewed by:
Judit Gervain, CNRS – Université Paris Descartes, France
Jan P. De Ruiter, Bielefeld University, Germany

***Correspondence:**
Imme Lammertink, Centre for Language Studies, Radboud University, Erasmusplein 1, 6525 HT Nijmegen, Netherlands
immelammertink@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *01 December 2014*
Accepted: *07 April 2015*
Published: *24 April 2015*

Citation:
Lammertink I, Casillas M, Benders T, Post B and Fikkert P (2015) Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. Front. Psychol. 6:495. doi: 10.3389/fpsyg.2015.00495

Introduction

Speakers in conversation take turns at talking (Sacks et al., 1974). The timing of speaker transitions is precise, usually exhibiting a 200 ms gap or a brief period of vocal overlap between turns (Stivers et al., 2009). Considering that it takes approximately 600 ms to initiate speech production (based on object naming; Levelt, 1989), addressees must anticipate when the current speaker's turn will end and must start planning their response well in advance to achieve minimal gap and minimal overlap transition timing (Levinson, 2013). This process requires the addressee to perform multiple tasks at once—decoding and interpreting the speech signal, plus formulating and articulating an appropriate response—all within the last few syllables of the ongoing turn (Levinson, 2013). Children, whose linguistic skills are still developing, have a hard time accomplishing these multiple tasks for turn-taking; it takes them several years before they master adult-like turn-taking behavior (age 6; Casillas et al., in press; Ervin-Tripp, 1979; Garvey, 1984).

Despite their late mastery of turn-taking, children begin taking turns (of a sort) in infancy. Caregivers respond to their 3–4-month-old infants' vocalizations, movements, and vegetative sounds as if they were "turns" in proto conversation (Bruner, 1975; Snow, 1977; Ginsburg and Kilbourne, 1988). Twelve-month-olds already understand conversational patterns well enough to expect speech (but not non-speech) sounds to provoke a verbal response from an

addressee (Thorgrímsson et al., 2015). One- and two-year-olds watching videos of conversation look anticipatorily to the upcoming responder at points of turn transition (Casillas and Frank, 2013). However, in their spontaneous turn-taking behavior, and in their predictions about upcoming turn boundaries, children are generally slower and less accurate compared to adults.

To anticipate upcoming turn structure accurately, children must learn to use predictive information in the ongoing speech signal. Recent experimental findings demonstrate that toddlers use both lexicosyntactic and prosodic information to predict upcoming speaker switches, but the relative importance of these information sources for prediction remains largely undetermined (Casillas and Frank, 2012, 2013; Keitel et al., 2013; Keitel and Daum, 2015). The current study investigates how Dutch and English toddlers weigh lexicosyntactic and prosodic[1] cues against one another in their online prediction of upcoming speaker switch.

Lexicosyntactic cues can provide critical information about upcoming speaker switches. For example, incomplete syntactic structures ("I'm making the…") hint that there is still more information to come, frequent multi-word sequences or strong semantic associations between words ("I need to brush my…") can strongly indicate what exact words will come next, and the word order of an utterance (interrogative vs. declarative) can help listeners predict how the current turn will finish (and what will happen in the next turn). Lexicosyntactic information appears to be critical for adult turn-end prediction: listeners anticipate turn-end timing more accurately when they can predict the exact words that will make up the rest of the turn (Magyari and de Ruiter, 2012). Speaker changes also almost always occur at points of lexicosyntactic completion in task-oriented dialogs (Dutch: Caspers, 2003; English: Ford and Thompson, 1996), and lexical and syntactic cues to questionhood (e.g., *wh*-words and subject-auxiliary inversion) occur early in the turn, thereby giving addressees more time to begin planning their response early (Bögels et al., 2014).

At least one previous study suggests that lexicosyntactic information is more important than prosodic information in adults' predictions about upcoming speaker changes. de Ruiter et al. (2006) asked participants to listen to fragments of speech and to press a button when they felt that the speaker's turn was coming to an end. Listeners achieved the same button press accuracy for normal speech (with full linguistic information) and intonationally flattened speech (with lexicosyntax, rhythm, and intensity, but no intonational information). In contrast, participants' accuracy significantly decreased for low-pass filtered speech (with full prosodic cues, but no lexicosyntactic information). The authors took this result as evidence that lexicosyntactic cues are primary, and possibly sufficient, for adult turn prediction, while prosodic cues play a less important role.

[1]The present study focuses on the role of intonation (one aspect of prosody) in children's prediction of upcoming speaker switches. We designed the stimuli to control for a few specific intonational contours. But, because we used a full, unfiltered linguistic signal, our intonational contours were accompanied by other prosodic cues (e.g., duration and intensity). For this reason we adopt the broader term "prosody" rather than the more narrow term "intonation."

Other work has characterized lexicosyntactic and prosodic cues as having qualitatively different functions for turn prediction. Under this view, lexicosyntax is particularly important in assessing whether a turn is complete and, by extension, whether it is ripe for a speaker switch. In natural speech, Dutch and English listeners rarely expect speaker switches when lexicosyntactic information is incomplete, no matter what intonational contour is used (Caspers, 2001; Wichmann and Caspers, 2001). But, when speakers have multi-utterance turns, and the addressee has to pass over several lexicosyntactically complete phrases before reaching the true turn-end, lexicosyntax alone does not provide sufficient information. Then prosody plays a critical role in listeners' ability to discriminate between *potential* completion points and *true* completion points. Turn-ends are often accompanied by prosodic cues such as boundary tones, increased syllable length, and post-turn silence (Ford and Thompson, 1996). Whether listeners expect a speaker change at lexicosyntactically complete points is largely dependent on the prosodic cues in the utterance (Caspers, 2001; Wichmann and Caspers, 2001).

The present paper addresses how children learn to use lexicosyntactic and prosodic cues in their prediction of upcoming turn structure. Generally speaking, children are sensitive to prosodic information before they become sensitive to lexicosyntactic information. Newborn infants use prosodic cues to distinguish their native language from other languages (Nazzi et al., 1998). Seven-month-olds can also use prosodic information to distinguish between words spoken with an angry, happy, or neutral voice (Grossmann et al., 2005). By 10 months, they can also use prosodic cues to segment the speech stream into smaller units (Gleitman and Wanner, 1982; Jusczyk, 1997; Christophe et al., 2008).

It is often assumed that children's sensitivity to prosodic information bootstraps their sensitivity to lexicosyntactic information (Morgan and Demuth, 1996; Christophe et al., 2008; Männel and Friederici, 2010). Newborns can discriminate categories of function words and content words on the basis of their different prosodic characteristics (Shi et al., 1999). Children show sensitivity to the word order of their native language as young as 7–8 months of age on the basis of word frequency and prosody (Höhle and Weissenborn, 2003; Gervain and Werker, 2013). Once children's knowledge of lexicosyntactic information becomes more detailed, they can access lexical and syntactic structures independently from the prosodic information available. For example, children start to recognize distinct function words at 11 months of age (Shi et al., 2006) and children at 12 months of age can use differences in word order to distinguish between questions and declaratives (Geffen and Mintz, 2014).

Given that sensitivity to prosodic cues precedes, or even bootstraps, sensitivity to lexicosyntactic cues, prosodic cues might have an early and primary role in children's predictions about upcoming speaker change. But recent studies have only found mixed evidence for this hypothesis. Casillas and Frank (2013) showed videos of conversation to children and adults, finding that children three and younger needed prosodic information to make above-chance anticipatory gaze switches to upcoming speakers in the video. In the same study, children three and older *did* show more gaze switches for lexical-only stimuli than

for prosody-only stimuli, but only for question-answer speaker switches: in conditions where lexical information was available, children made more anticipatory gaze switches after hearing questions than non-questions. Their results suggest an early, more global role of prosody in turn prediction and a later, question-specific role of lexicosyntax. Importantly, the stimuli in their experiment were phonetically manipulated to control for linguistic information, e.g., using speech that was low-pass filtered, intonationally flattened, duration-controlled, and multi-layered (but see also Casillas and Frank, 2012). Children do not often hear this kind of phonetically controlled speech in their natural language environment.

Keitel et al. (2013); Keitel and Daum (2015) performed a similar study, showing children videos of conversation and using children's age (rather than phonetic manipulation) to control for the availability of lexicosyntactic cues; they tested both pre-verbal (6- and 12-month-old) and verbal (24- and 36-month-old) children. To test for the role of intonation, half of the videos featured pitch-flattened speech and the other half featured a full linguistic signal. Children only made above-chance anticipatory gaze switches to the upcoming responder at 36 months—considerably later than what Casillas and Frank (2013) found—and anticipated speaker changes less often when intonational contours were removed (but only at 36 months). In contrast, adults' turn predictions were unaffected by the lack of intonational contours. The findings indicate that intonation may be useful for children's turn prediction, but only at age three and up, and not for adults. But, again, the primary linguistic control in the stimuli depended on phonetic manipulation of the speech signal. Thus, the results of these prior studies—Casillas and Frank (2013) and Keitel et al. (2013); Keitel and Daum (2015)—are based on a comparison between natural (full signal) and non-natural (phonetically manipulated) stimuli.

Unnatural speech is noticeable to children, and more generally changes the way listeners process linguistic information. Twelve- and 36-month-olds prefer speech sounds to non-speech (motor) sounds while watching videos of conversation (Bakker et al., 2011). If children in the studies mentioned above interpreted the manipulated speech as degraded or even as non-speech, they might have processed the lexicosyntactic and prosodic information differently than they do in everyday interactions. Even for adults, acoustically unusual stimuli, such as synthetic speech, can cause significant processing costs (Pisoni, 1981).

The current study is designed to assess the relative and the individual contributions of prosody and lexicosyntax for turn structure prediction while using the full speech signal (unfiltered speech with both lexicosyntactic and prosodic cues present). We used a full speech signal so that we could test children's use of linguistic cues for speaker-switch prediction with stimuli that resemble speech in their natural environment—stimuli without any phonetic filtering or resynthesis. Participants watched eight videos of short, scripted conversation. For a subset of the utterances in each conversation, we controlled for the presence of lexicosyntactic and prosodic cues to turn completion by cross-splicing snippets from multiple sentence recordings (see Section Stimulus Preparation). In one condition, both lexicosyntax and prosody signaled an upcoming speaker switch (a fully complete turn). In the opposite condition, neither cue signaled an upcoming speaker switch (a fully incomplete turn). In two more conditions, lexicosyntax, and prosody were pitted against each other to test for their relative primacy (i.e., complete lexicosyntax with incomplete prosody *or* incomplete lexicosyntax with complete prosody). We expected that young children would rely more on prosodic cues in their prediction of upcoming turn structure, given their early acquisition of basic prosodic knowledge.

Following recent work, we measured children's predictions about upcoming turn structure by tracking their eye movements while they watched videos of dyadic conversation between puppets. In line with previous studies investigating children's anticipation of turn structure (Casillas and Frank, 2013; Keitel and Daum, 2015), we used puppet dyads to capture children's attention while also conveniently removing the non-verbal cues to turn taking that often appear at turn boundaries (e.g., gaze and gesture; Rossano et al., 2009; Stivers and Rossano, 2010). The absence of non-verbal cues enabled us to focus on the role of linguistic cues.

Eye tracking is a natural and passive measure of attention, but provides an online measure of children's predictive processing during conversation (Casillas and Frank, 2012, 2013; Keitel et al., 2013; Keitel and Daum, 2015). Prior work has shown that, compared to explicit measures of turn-end prediction (e.g., button-press; de Ruiter et al., 2006), anticipatory eye movements from the prior to the next speaker tend to occur quite late at points of speaker transition (children: Casillas and Frank, 2012, 2013; Keitel et al., 2013; Keitel and Daum, 2015; and adults: Tice and Henetz, 2011; Hirvenkari et al., 2013; but also see Holler and Kendrick, 2015 for earlier switching in adults). Eye-tracking measures therefore do not target turn-end prediction the same way that button press measures do. Instead, they appear to index the prediction of upcoming turn transitions and the onset of an upcoming response, both of which are affected by linguistic material present in the pre-transition turn (e.g., question vs. non-question, prosodic information, etc…). We track participants' anticipatory eye movements to upcoming speakers as a natural measure of their predictions about upcoming turn structure.

We sampled from two linguistic populations to test the robustness of our findings: Dutch (Experiment 1) and British English (Experiment 2). Dutch and English use similar linguistic structures to form simple declaratives and polar interrogatives; in both languages the subject precedes the verb in declarative utterances, whereas interrogative utterances are created by subject-verb inversion (Dryer, 2013). The prototypical intonation pattern for polar questions in both languages also features a final rise[2] (e.g., Dutch: Haan, 2002; English: Grabe and Post, 2004).

Experiment 1

Materials and Methods

Participants

Thirty-three native Dutch-speaking 2.5-year-olds participated in the experiment. Of these, twelve were excluded because of

[2]Final-rising contours are considered prototypical for polar questions, but there are also many other intonation contours used with polar questions in spontaneous speech (see, e.g., Couper-Kuhlen, 2012).

equipment error (1) or inattention to the screen during the experiment (11; see Section Data Pre-Processing). As a result, 21 toddlers were included in the final analysis (Female = 13, mean age = 29 months, range = 24–33 months). Sixteen adult participants (native Dutch speakers, Female = 15, mean age = 23 years) participated as a control group. No hearing or vision problems were reported. Ethical approval for the study was obtained from the *Ethiek commissie faculteit der Sociale Wetenschappen (ECSW)* at Radboud University in Nijmegen.

Apparatus

We recruited and tested toddlers through the Baby Research Center (BRC) in Nijmegen, The Netherlands. The data were obtained with a 17-inch Tobii 1750 eye-tracker (Tobii Technology AB; binocular infrared light reflection, 50 Hz sampling frequency, accuracy range: 0.5° to 1°, recovery <100 ms). Eye-tracker calibration and stimulus presentation were controlled by ClearView 2.7.1 software. Audio speakers were placed at either side of the screen, hidden from participant view. Participants sat approximately 60 cm from the monitor, with toddlers sitting on their parent's lap.

Procedure

Each session began with a 9-point infant-friendly calibration procedure. Data collection started when good calibration for both eyes was obtained for at least five locations on the screen (every corner and the center). Children then watched eight 30-s videos of conversation between two puppets. Before each conversation, the experimenter displayed an animated smiley face on the screen until children's gaze returned to the center. After every two conversations, participants saw a 4–9-s animated filler video (a train, a skating dog, and a running chick). The experiment took 5 min in total. Two versions of the experiment were created, with conversation videos ordered differently in each. In both versions, the same pair of puppets was shown, at most, twice in a row. Participants were randomly assigned to one of the two versions.

Audio Stimuli

Target utterances

We created four types of target utterances by controlling for lexicosyntactic and prosodic cues to turn completion (**Table 1**). At the point of syntactic completion (or incompletion) for each target utterance we inserted 500 ms of silence ("[…]" in **Table 1**). Participants could then make a prediction, depending on the linguistic information up to that point, about whether the same speaker would continue or whether the addressee would respond. We measured participants' anticipatory gaze to the addressee around these 500 ms silent windows.

The utterances with cues to turn completeness featured polar interrogative syntax (+SYN), a polar interrogative pitch contour (+PROS; a high, final rise in Standard Dutch; Haan, 2002), or both. The utterances with cues to turn incompleteness featured incomplete declarative syntax (−SYN), incomplete non-interrogative pitch contours (−PROS), or both. The incomplete non-interrogative pitch contours were deemed "incomplete" because they lacked boundary tones at the onset of the inserted 500 ms silence.

By this design, fully complete utterances were both lexicosyntactically and prosodically complete, and took the form of polar interrogatives with a final rise pitch contour, followed by 500 ms of silence (e.g., *Shall we swim together?* […]). Meanwhile, fully incomplete utterances were both lexicosyntactically and prosodically incomplete at the onset of the 500 ms silence. These fully incomplete utterances took the form of declarative sentences that had been split into two parts by 500 ms of silence; at the onset of the silence (where we measured participants' anticipatory gaze),

TABLE 1 | (A) Examples of target utterances in the four conditions. (B) Example of a conversation with the four target utterances embedded in six filler utterances.

(A) Target utterances			(B) Conversation	
Condition	**Cues**	**Example**	**Speaker**	**Example**
			A	I think I'll go swimming today
(1) Fully incomplete	−SYN −PROS	Today is a beautiful+ […] +day for a swim	B (1)	Today is a beautiful+ […] +day for a swim
			B	And I have a new swimsuit
(2) Incomplete syntax	−SYN +PROS	It's made especially for? […] Swimming in the ocean	B (2)	It's made especially for? […] Swimming in the ocean
			A	Wow, you should try it out then
			A	I bet you bought a really nice one
(3) Incomplete prosody	+SYN −PROS	Do you enjoy swimming+ […]	B (3)	Do you enjoy swimming+ […]
			A	Yes, I like to swim a lot
(4) Fully complete	+SYN +PROS	Shall we swim together? […]	B (4)	Shall we swim together? […]
			B	That would be really fun

Each utterance is marked as syntactically complete (+SYN) or incomplete (−SYN), and prosodically complete (+PROS) or incomplete (−PROS). The symbol "?" indicates a complete, polar interrogative pitch contour, "+" indicates the lack of a boundary tone, and "[…]" indicates a 500 ms silence. Examples are taken from the British English stimuli used in Experiment 2.

the in-progress utterance was lexicosyntactically incomplete and had no final boundary tone (e.g., *Today is a beautiful*+ [. . .] +*day for a swim*).

The two other target utterance types were only partially complete. For example, utterances that were prosodically complete but lexicosyntactically incomplete took the form of declarative sentences that had been split into two parts by 500 ms of silence; at the onset of the silence, the in-progress utterance was lexicosyntactically incomplete but prosodically complete, with a final rise pitch contour (e.g., *It's made especially for?* [. . .] *swimming in the ocean*). Meanwhile, utterances that were lexicosyntactically complete but prosodically incomplete, took the form of complete polar interrogatives that lacked a final boundary tone at the onset of the 500 ms of silence (e.g., *Do you enjoy swimming*+ [. . .]). **Table 1** gives an example conversation that demonstrates the placement of the 500 ms silences for each utterance type.

With this design, all lexicosyntactically complete utterances were interrogative and all lexicosyntactically incomplete utterances were declarative. In designing the utterance types this way, we created a maximal contrast in participants' expectations about an upcoming turn switch between the *fully complete* and *fully incomplete* utterances. Questions naturally project an answer in the next turn, and so observers could reliably expect a turn transition after hearing a question (Casillas and Frank, 2012, 2013). Declaratives do not necessarily project a turn transition, and so observers' expectations after declaratives are much weaker. We sought to create a maximal difference in the fully complete and incomplete conditions because they served as the baselines for our primary conditions of interest: the partially complete conditions (incomplete syntax and incomplete prosody).

We could have instead tried to keep word order the same across the complete and incomplete lexicosyntactic conditions, but this would have created other problems. For example, using interrogative word order for all utterance types would have signaled turn transition early on in the utterance for all sentences, yielding ambiguous and unnatural sentences in the lexicosyntactically incomplete condition ("Would you like a+"). Using declarative word order in both conditions could have possibly worked; declarative polar questions *do* occur in spontaneous Dutch and English (Gunlogson, 2001; Englert, 2010). But declarative polar questions are primarily used for the initiation of repair or for confirmation requests, whereas interrogative polar questions are primarily used for requesting information (Englert, 2010). Thus, even if we used declarative word order for all utterance types, the speech acts would still differ across types. Additionally, to use declarative polar questions, we would need to generate the required contexts for declarative questioning into the scripts (e.g., potential mishearing/misunderstanding), thereby introducing further variation across conversations. Considering these issues together, we decided to use interrogative polar questions for lexicosyntactically complete conditions and unfinished declaratives (at the onset of the 500 ms silence) for lexicosyntactically incomplete conditions. There were eight target sentences in each conversation, resulting in 32 total target sentences. Each of the four conditions for target sentences is described below.

Conversation design

The targets were embedded in eight 30-s scripted conversations about topics familiar to 2.5-year-olds (rabbits, snowmen, swimming, birthday parties, and bicycles; Zink and Lejaegere, 2003). Every conversation had six filler and four target utterances, including one target utterance from each type (**Table 1B**). Targets and filler utterances were separated by 500 ms of silence.

The order of the target utterances within the eight conversations was counterbalanced. Target utterances were equally divided between the two speakers across the eight conversations of the experiment. After *fully incomplete* and *incomplete syntax* target utterances, no turn transition occurred following the 500 ms of silence; the current speaker always completed her turn. After *incomplete prosody* and *fully complete* target utterances, target turns were followed by 500 ms and then a change in speakership 50% of the time. Each conversation contained from five to seven turn transitions.

Stimulus preparation

The audio stimuli were recorded in a sound-attenuated booth by two female native speakers of Standard Dutch. The audio for each experiment was collected over two recording sessions. In the first session, both speakers were recorded simultaneously while they acted out the eight dialogs together, three times each. Speakers were asked to read each conversation in an infant-directed register. The filler utterances were then extracted from the best recording of each conversation and then set aside for use in the final stimuli. In the second recording session, speakers were recorded individually as they read an additional set of recording utterances aloud. The additional recording sentences were designed to elicit sub-parts of the target sentences—subparts that could then be spliced together to create the final target utterances (see below). In the second recording, speakers matched their pitch, speaking rate, and affect to the first recording by listening to the first-session conversations over a pair of headphones. The final target utterances were then spliced together from these second-session utterances, and then the conversations were spliced together from a combination of the filler and target utterances.

We composed each target utterance from three or four parts: an initial part, a prosody part, a silence, and (for lexicosyntactically incomplete utterances) a completion part (**Figure 1**). Each part derived from a separate recording utterance (from the second recording session). The parts were then spliced together to obtain the final set of target utterances (Praat; Boersma and Weenink, 2012).

The "initial part" of the target utterance was two words long, with an utterance-initial non-interrogative prosodic contour. For example, the "*I've seen*" in "*I've see carrots behind the swings*" was extracted from the recording sentence "***I've seen*** *caramel*" (**Figure 1**).

The "prosody part" was also two words long. For prosodically incomplete target utterances, there was no prosodic boundary at the end of the second word. For example, "*carrots behind*" was extracted from the recording sentence "*I've seen* ***carrots behind*** *the broccoli for weeks*". In these recording sentences, the utterance always continued beyond the splicing point to ensure that

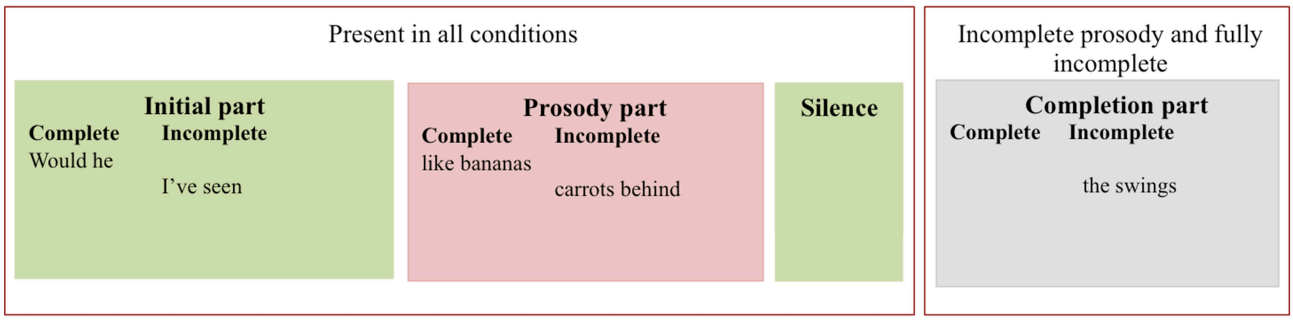


FIGURE 1 | Division of two target utterances into their subparts. The initial part (green), the prosody part (red), and the silence (green) were present in all target utterances. The completion part (gray) was only present in lexicosyntactically incomplete target utterances.

there was no intonational phrase boundary at the end of the two-word prosodic part. On the other hand, for prosodically complete utterances, the two-word prosody part had a complete, interrogative prosodic contour. For example, "*like bananas*" was extracted from the recording sentence "*Said he:* '***like bananas?***'" (**Figure 1**).

The prosody part was followed by 500 ms of silence. Although 500 ms is somewhat long for an inter-turn gap in adult conversation (Stivers et al., 2009), it closely resembles the median response latency for children in interaction with their parents (549 ms for children's responses at 2;4–2;5; Casillas et al., in press) and it is much shorter than their median response latency with their peers (900 ms for children's responses at 2;10–3;3; Garvey and Berninger, 1981). A pause of 500 ms also gives participants (especially the children) substantial time to process the lexicosyntactic and prosodic information in the utterance preceding a turn transition. The 500 ms window also allowed reliable measurement of children's anticipatory eye movements because toddlers need at least 300 ms to plan a shift in gaze (Fernald et al., 2001).

Finally, the completion part (only present in the lexicosyntactically incomplete utterances) contained between one and five words that syntactically completed the pre-silence portion. For example, "*the swings*" was extracted from "*I've seen carrots behind* ***the swings***" (**Figure 1**).

To avoid audibly mismatched co-articulation, we matched the place of articulation for phonemes at splicing boundaries. For example, the initial part "*I've seen*" was followed by a/k/in the recording sentence to match the initial/k/of the prosody part "*carrots behind*." That way, when spliced together, "*I've seen*" + "*carrots behind*" had no conflicting co-articulatory cues. Similarly, we avoided co-articulatory cues to upcoming speech by controlling the phonemes immediately following incomplete prosody parts. For example, "*carrots behind*" was followed by an/ð/("the") in the recording sentence. Because the/d/in "*behind*" and the/ð/in "*the*" approximately match in place of articulation, there is no co-articulation to cue further upcoming speech. Alternately, the prosody part was followed by a phoneme with a neutral place of articulation (/ʔ/or/h/), matched for the 500 ms silence.

We also controlled for primary stress in the two-word initial parts that had interrogative word order (*fully complete* and *incomplete prosody)*. Though the primary prosodic cue for polar interrogatives is a final high rise, they also often have high fundamental frequency at the start of the utterance (Haan, 2002). To counteract this and to also prevent the presence of prosodic boundary tones at points of intended prosodic incompleteness, we asked speakers to put emphasis on words that came late in the utterance, thereby avoiding stress placement at the start of the utterance or at the intended splicing points.

Stimulus pre-testing

We verified the status of our utterances as lexicosyntactically complete/incomplete with a web-based experiment using a written version of the utterances. Fourteen participants (Female = 7, mean age = 23.8 years old, native Dutch speakers) read and judged the completeness of the thirty-two (16 −SYN and 16 +SYN) target sentences up to the point of the inserted 500 ms silence (Qualtrics Software Version 55939, 2014[3]). All target sentences were found to be complete or incomplete, as intended, by more than 75% of the participants.

We verified our manipulation of prosodic completeness with a listening experiment conducted in Praat (Boersma and Weenink, 2012). Twelve participants (Female = 10, mean age = 24 years, native Dutch speakers) heard low-pass filtered versions of the target utterances (300 Hz and 50 Hz Hanning window), and were asked to judge whether each one was a question or not. Low-pass filtering removes segmental information so that only prosodic information remains. Each target utterance was presented twice, with the order of utterances fully randomized. Eleven (34%; five complete and six incomplete) targets were judged differently than intended (e.g., judged as an interrogative contour when it should have been non-interrogative, or vice versa) in more that 25% of the judgments. These ambiguous prosodic contours were therefore taken into account during data analysis and interpretation.

Video Stimuli

Two pairs of puppets were used to create the stimulus videos. To match the puppet videos to the audio stimuli as closely as

[3]Qualtrics (2014), Provo, Utah, USA. Available from: http://www.qualtrics.com (Version 55939).

possible, two puppeteers listened to the dialogs and simultaneously moved the puppet mouths during video recording. The puppeteers aimed to complete an open-close mouth movement for each syllable in the recording. With the exception of mouth movements, the puppets were immobile. We then combined the puppet video recordings with the audio stimuli, maximizing the quality of sound and speech alignment in Adobe Premiere Pro video editing software.

Data Pre-Processing

Before analyzing children's anticipatory gaze switches, the raw data set was pre-processed to remove unreliable tracker output and to prepare gaze measurements for the main gaze-switch analysis. We only counted participants' gaze measurements when the Tobii output marked the look as valid in at least one eye. Trials were excluded when a participant attended to the screen for less than 75% of the total trial duration. If this happened for more than four trials, the participant's data was completely excluded from further analysis because of a general inattention to the stimuli. In total, eleven toddlers (33%) were completely excluded by this criterion. No adults were completely excluded. From the remaining participants (21 toddlers and 16 adults), 27 trials (2.5%) were excluded in total (toddlers: 21, adults: 6). The final dataset contained gaze data for 1056 trials.

Our main question was how toddlers use linguistic cues in their *prediction* of upcoming speaker changes, so we only analyzed gaze switches that were initiated before children could have reacted to speaker continuation/speaker switch. We used an algorithm for switch identification developed by Casillas and Frank (under review). According to this three-step checklist, switches are anticipatory if they fulfill the following criteria: (1) a participant fixates on the prior speaker for at least 100 ms at the end of the prior turn, (2) sometime thereafter the participant switches to fixate on the upcoming speaker for at least 100 ms and, (3) a gaze shift is initiated within the first 300 ms of the response turn for toddlers (Fernald et al., 2001), or 200 ms for adults[4].

Random gaze switches between speakers can sometimes, by chance, conform to these three criteria, and could therefore be mistakenly categorized as "true" gaze switches. Therefore, we estimated and corrected for participants' baseline random anticipatory looking behavior. Again, algorithmic details were borrowed from Casillas and Frank (under review). We ran each participant's actual eye-tracking data through the exact same switch identification algorithm (described above), but this time with 100 randomly-shuffled versions of the original turn-transitions in the videos (**Figure 2**). The idea was that, if we assume as our null hypothesis that children's switching behavior is random, their rates of anticipatory switching should be the same no matter where we place our analysis windows (at real turn transitions vs. anywhere else in the stimulus). We therefore made 100 versions of the original analysis windows in which the original analysis windows for each stimulus were distributed randomly between its start and end time (**Figure 2**). Then, using the three-step algorithm described above, we determined whether the participant made an anticipatory switch or not for each turn transition in each randomly-shuffled version. This procedure was repeated 100 times. Then we averaged the results to get a single baseline estimate of random switching for each target turn transition for each participant. We then obtained corrected anticipatory gaze switch values by subtracting the random anticipatory gaze switch value from the original gaze switch value for each turn transition for each participant. These corrected anticipatory switch values were then used in all statistical analyses (see also **Figures 3**, **5**).

[4]The children in the current experiment are older than those in Casillas and Frank (under review; ages 1–6), so we adopted a shorter window (300ms, not 333ms) for the assumed time needed for children to plan a gaze shift (Fernald et al., 2001).

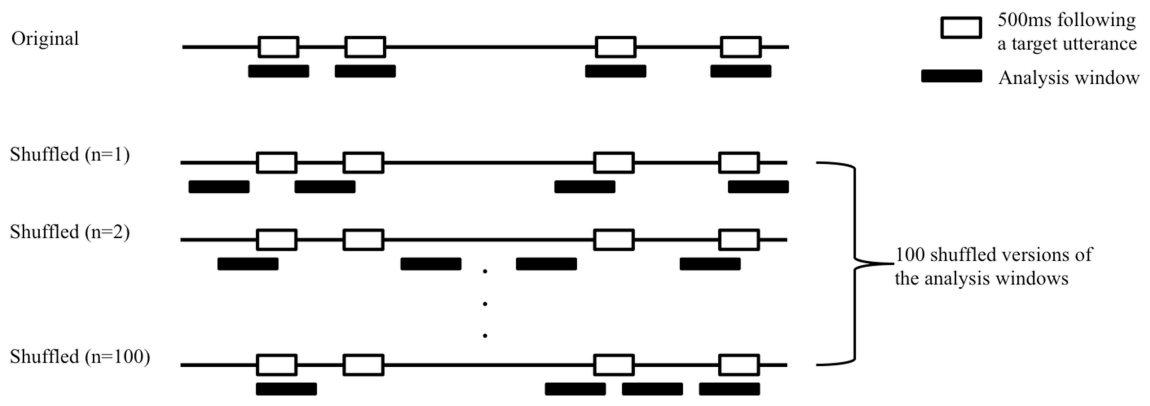


FIGURE 2 | Shuffling procedure for estimating random baseline anticipatory switches. We created 100 randomly shuffled versions of the original analysis window placement. The horizontal line represents the duration of one example dialog. The white boxes on the line represent the four 500 ms silences after potential target turn transitions (32 in total across the 8 dialogs; See **Table 1**). The black boxes beneath the line represent the four analysis windows for each target 500 ms silence. In the original version, the analysis windows are centered on the target 500 ms silences. In the shuffled versions, the analysis windows are randomly redistributed over the duration of the turn. We then re-ran the gaze identification algorithm, pairing the eye tracker data with each randomly shuffled version to estimate the baseline probability of making an anticipatory shift when the analysis windows are uncoupled from the actual target 500 ms silences.

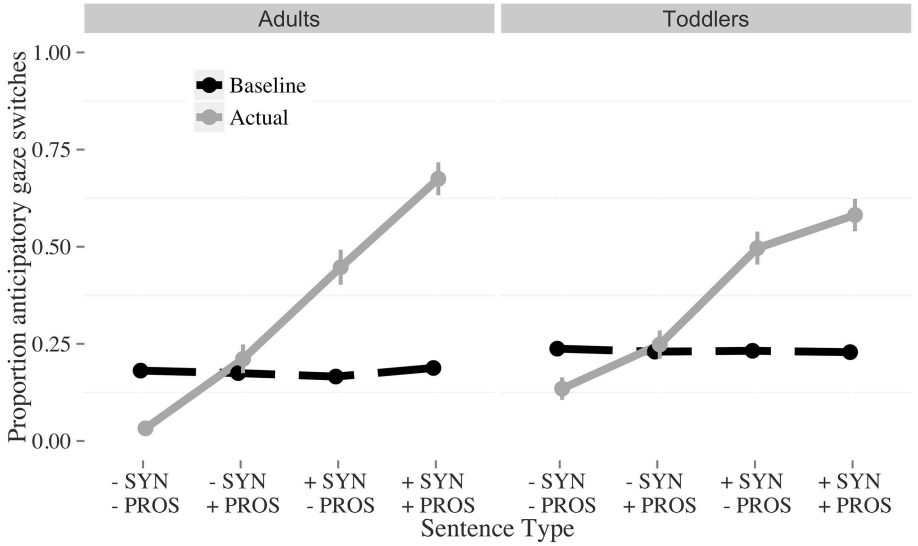


FIGURE 3 | Proportion of baseline (dashed) and actual anticipatory (solid) gaze switches to the answerer, by condition and age. The conditions were: *Fully incomplete* (−SYN −PROS); *Incomplete syntax* (−SYN +PROS); *Incomplete prosody* (+SYN −PROS); *Fully complete* (+SYN +PROS). The vertical bars indicate the standard error of the mean.

Results

The complete pre-processed dataset (toddlers and adults together) was analyzed using linear mixed effect models (lme4; Bates et al., 2012) in the statistical programming language R (R Development Core Team, 2013). Significance of the predictors was evaluated by using the obtained z-score as a t-statistic ($|t|>1.96$ is significant at $\alpha = 0.05$).

Pre-Analysis: Random Anticipatory Looking

The original anticipatory gaze switches and the random baseline anticipatory gaze switches are visualized in **Figure 3**. Participants switch less than would be expected by chance in the *fully incomplete* (−SYN, −PROS) condition, at chance level for the *incomplete syntax* (−SYN, +PROS) condition, and above chance for both the *incomplete prosody* (+SYN, −PROS) and *fully complete* (+SYN, +PROS) conditions. This pattern suggests that participants use both lexicosyntactic and prosodic cues for turn-projection: When both cues are incomplete, participants do not expect a speaker change, whereas when both cues are complete, they do. When the cues are pitted against each other, listeners weigh lexicosyntactic over prosodic cues.

Lexicosyntactic and Prosodic cues

In order to assess the effects of linguistic cue and participant age, we first fit a model to participants' baseline-corrected anticipatory switches (1056 observations; $N = 37$; **Table 2**). All targets ($N = 32$) were included in the model. Recall that the prosodic pre-test (Section Conversation Design) showed that eleven targets had ambiguous prosodic contours. A model including these ambiguous targets did not reveal qualitatively different results compared to a model excluding these targets. Therefore, the final model included all targets. The dependent variable was participants' baseline-corrected anticipatory gaze switches. Predictor variables included *syntactic completeness* (incomplete vs. complete), *prosodic completeness* (incomplete vs. complete) and *age* (toddler vs. adult). The predictor variables were contrast-coded (**Table 2**) and the intercept was allowed to vary by subject and item.

TABLE 2 | Outcomes from the linear mixed effects model including both subject groups (Dutch toddlers and adults; Number of observations: 1056; *N* = 37).

Predictor	Contrast coding	*β*	*t (z)*	*p*
Intercept		0.15	4.77	
Syntactic completeness	Incomplete (−1) Complete (1)	0.20	6.67	<0.001
Prosodic completeness	Incomplete (−1) Complete (1)	0.075	2.56	<0.05
Age	Toddler (−1) Adult (1)	0.019	1.28	
Syntactic completeness × Prosodic completeness		−0.0015	−0.051	
Syntactic completeness × Age		0.024	2.03	0.05
Prosodic completeness × Age		0.022	1.82	
Syntactic completeness × Prosodic completeness × Age		0.0065	0.54	

The amount of linguistic information consistent with turn completion affected participants' anticipatory switching. Model coefficients show three significant effects in the anticipatory gaze data. First, the proportion of anticipatory gaze switches was larger for the lexicosyntactically complete vs. incomplete targets ($\beta = 0.20$, $z = 6.67$, $p < 0.001$). Second, more anticipatory gaze switches were made for complete prosodic contours than for incomplete prosodic contours ($\beta = 0.075$, $z = 2.56$, $p < 0.05$). Third, there was an interaction between *syntactic completeness*

and *age* (β = 0.024, $z = 2.03$, $p = 0.05$). No other coefficients reached significance.

Visual inspection of the data (**Figure 4**) suggests that interaction between syntactic completeness and age comes from the *fully complete* condition, in which toddlers and adults differed in their overall number of anticipatory switches (adults switch more than toddlers do). We fitted a model restricted to the syntactically complete conditions (*fully complete* and *incomplete prosody,* 528 observations, $N = 37$, **Table 3**) to test this hypothesis. No significant effect of *age* was found. Therefore, this explanation for the interaction was not verified in the statistical model.

Relative Weight of Lexicosyntactic and Prosodic Cues

A second model was fit to test the relative weight of lexicosyntactic and prosodic cues. We restricted the data to the two partially complete conditions (*incomplete syntax* and *incomplete prosody*). The predictor variables were *condition* (incomplete syntax vs. incomplete prosody) and *age* (toddler vs. adult). Again, the intercept was allowed to vary by subject and item.

The model showed that participants weighed lexicosyntactic cues over prosodic cues. Participants made more anticipatory gaze switches after utterances with complete syntax (*incomplete prosody*) compared to utterances with complete prosody (*incomplete syntax*; β = 0.12, $z = 2.62$, $p < 0.05$; **Table 4**). No other predictors reached significance.

Speaker Change or Speaker Continuation

Recall that lexicosyntactically complete targets were followed by a change in speakership 50% of the time. Ideally there should be no difference between cases of speaker change and speaker continuation. If there were a difference, it could indicate that participants' gaze switches were triggered by the mouth movement of the responding or continuing puppet rather than by the participants' predictions alone.

We ran an additional analysis to test whether anticipatory gaze switches were influenced by speaker change in the lexicosyntactically complete conditions (*fully complete* and *incomplete prosody*), with speaker continuation (-1) and speaker change (1) contrast-coded. The analysis was restricted to the lexicosyntactically complete conditions since syntactically incomplete targets were always followed by continuation of the same speaker.

The model (528 observations, $N = 37$) revealed a significant effect of *speaker change*. Participants made more anticipatory gaze switches when the target was followed by a speaker change compared to a speaker continuation (β = 0.18, $z = 5.26$, $p < 0.001$; **Table 5**). No other coefficients reached significance.

A closer look at the video stimuli indeed showed that sound and mouth movement were not adequately aligned in almost half of the syntactically complete target utterances: In nine of

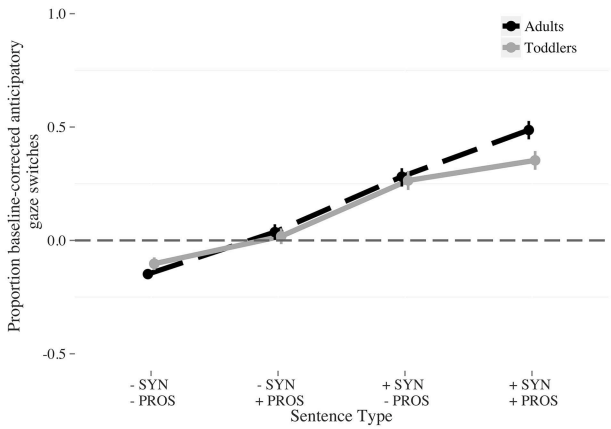


FIGURE 4 | Proportion baseline-corrected anticipatory switches to the answerer by condition and age (Adults: dashed; Toddlers: solid). The conditions were: *Fully incomplete* (−SYN −PROS); *Incomplete syntax* (−SYN +PROS); *Incomplete prosody* (+SYN −PROS); *Fully complete* (+SYN +PROS). The vertical bars indicate the standard error of the mean.

TABLE 3 | Result from the linear mixed effects models for Dutch toddlers and adults by lexicosyntactic complete targets (528 observations, *N* = 37).

Predictor	Contrast coding	β	t(z)	p
Intercept		0.34	5.71	
Prosodic completeness	Incomplete (−1) Complete (1)	0.074	1.29	
Age	Toddler (−1) Adult (1)	0.044	1.70	
Age × Prosodic completeness		0.028	1.52	

TABLE 4 | Outcomes from the linear mixed effects model of the two partially complete conditions for both subject groups (*incomplete syntax*, *incomplete prosody*, Number of observations: 528; *N* = 37).

Predictor	Contrast coding	β	t (z)	p
Intercept		0.15	3.14	
Condition	Incomplete syntax (−1) Incomplete prosody (1)	0.12	2.62	<0.05
Age	Toddler (−1) Adult (1)	0.012	0.57	
Condition × Age		0.0026	0.15	

TABLE 5 | Results from the linear mixed effect models for Dutch toddlers and adults in the lexicosyntactically complete conditions, including the predictor variable speaker change (528 observations, *N* = 37).

Predictor	Contrast coding	β	t (z)	p
Intercept		0.34	8.88	
Prosodic completeness	Incomplete (−1) Complete (1)	0.074	2.16	<0.05
Age	Toddler (−1) Adult (1)	0.044	1.69	
Speaker change	No (−1) Yes (1)	0.18	5.26	<0.001
Prosodic completeness × Age		0.029	1.54	
Prosodic completeness × Speaker change		−0.010	−0.31	
Age × Speaker changes		0.022	1.17	
Prosodic completeness × Age × Speaker change		−0.0078	−0.42	

the sixteen lexicosyntactically complete target utterances, mouth movement preceded the onset of the audio signal by more than a few milliseconds. This early mouth movement could have triggered participants' gaze shifts toward the moving puppet, regardless of the linguistic content available.

Additionally, because the prior speaker *always* continued after the silence for lexicosyntactically incomplete targets, but only continued 50% of the time after lexicosyntactically complete targets, there was a statistical bias in the stimuli that could have caused participants to make fewer anticipatory gaze switches for lexicosyntactically incomplete targets. If so, participants would have to learn this statistical bias during the course of the experiment; it should only be present at the end of the experiment. We fit two linear mixed effect models to (a) data from the first two trials (268 observations, $N = 36$) and (b) data from the last two trials (248 observations, $N = 34$). In both models, the main effect of lexicosyntactic completeness was present (First two trials: $\beta = 0.146$; $z = 3.052$, $p < 0.01$; Last two trials: $\beta = 0.199$, $z = 4.885$, $p < 0.0001$, **Table 6**). The results therefore do not support statistical learning as a basis for the effects of lexicosyntactic completeness.

Discussion

Both Dutch toddlers and adults used lexicosyntactic and prosodic cues in their anticipation of upcoming speaker changes. Participants made the least anticipatory gaze switches when both cues signaled an incomplete turn. The most anticipatory gaze switches were made when both cues signaled a complete turn. When the lexicosyntactic and prosodic cues were pitted against each other (*incomplete syntax* and *incomplete prosody*), listeners weighed lexicosyntactic over prosodic cues.

The advantage for lexicosyntactic over prosodic cues in turn-projection is consistent with prior work on adult turn-taking (Caspers, 2001; de Ruiter et al., 2006), but was unexpected for toddlers. Recent work on children's use of prosodic and lexicosyntactic cues in predicting upcoming turn structure found an early global advantage for prosodic over lexicosyntactic cues in 1- and 2-year-olds (Casillas and Frank, 2013). An early advantage for prosodic cues would have also been consistent with the general pattern in language acquisition that sensitivity to prosodic cues precedes sensitivity to lexicosyntactic cues (Nazzi et al., 1998; Christophe et al., 2008).

Before accepting the hypothesis that 2.5-year-old toddlers weigh lexicosyntactic over prosodic cues in their turn-projection, alternative explanations need to be explored. A first explanation relates to the reliability of the prosodic contours in the stimuli. Recall that 11 of the 32 prosodic contours were ambiguous in whether they signaled interrogativity (completeness) or not; pre-test participants classified these 11 contours incorrectly at least 25% of the time. The results of the main experiment did not qualitatively shift when these ambiguous prosodic contours were included (Section Lexicosyntactic and Prosodic Cues), but their presence could have affected overall task performance. For example, toddlers may have noticed that the prosodic contours were strange or unclear and therefore unconsciously shifted their attention away from the prosodic cues in favor of the (less ambiguous) lexicosyntactic cues.

Another alternative explanation for toddlers' reliance on lexicosyntactic cues is that the puppet movements gave unintentional cues to turn hold or turn transition. *Post-hoc* analyses revealed that participants made more anticipatory gaze switches when lexicosyntactically complete turns were followed by a change in speakership compared to when they were followed by a continuation of the same speaker. We found that non-verbal cues (e.g., opening mouth, movements) preceded the onset of the acoustic signal in 9 of the 16 syntactically complete target utterances. These early non-verbal cues could have enhanced the effect of lexicosyntactic completeness, because early visual cues to speaker change were available in some of the lexicosyntactically complete target utterances, while lexicosyntactically incomplete target utterances were never followed with visual cues to speaker change (the same speaker always continued; **Table 1**).

Despite these methodological issues, the results from Experiment 1 still suggest that lexicosyntactic cues are weighed over prosodic ones in children's prediction of upcoming turn structure. To test the robustness (non-language specificity) of

TABLE 6 | Outcomes from the main linear mixed effects model for (A): first two trials of the experiment and (B): last two trials of the experiment (Dutch toddlers and adults).

	Dutch					
	(A). First two trials 268 observations, $N = 36$			(B). Last two trials 248 observations, $N = 33$		
	β	*t (z)*	*p*	*β*	*t (z)*	*p*
Intercept	0.114	2.364		0.0984	2.225	
Syntactic completeness	0.146	3.052	<0.01	0.199	4.885	<0.0001
Prosodic completeness	0.0382	0.798		0.0847	2.080	<0.05
Age	0.0301	1.130		0.000563	0.020	
Syntactic completeness × Prosodic completeness	−0.701	−1.465		0.0242	0.595	
Syntactic completeness × Age	0.0224	0.866		−0.0131	−0.575	
Prosodic completeness × Age	0.0192	0.744		0.0115	0.503	
Syntactic completeness × Prosodic completeness × Age	−0.679			0.0335	1.470	

these findings, we conducted a second experiment with British English toddlers and adults.

Experiment 2

Experiment 2 tested how English-speaking toddlers weigh prosodic and lexicosyntactic cues for upcoming turn structure prediction. Diverging slightly from Experiment 1, the recording and splicing for the target utterances in Experiment 2 used an extra criterion: the "prosody part" of the target utterances contained at least four syllables (only two were used in Experiment 1; see Section Stimulus Preparation). By extending the prosodic contour over more syllables, we gave the listener more time to perceive the contour being used. We derived the criterion of "four syllables" from the Dutch pre-test for prosodic completeness; most of the errors were made on *prosody parts* with fewer than four syllables. As in Experiment 1, participants' eye movements were recorded as they watched eight videos of dyadic puppet conversation.

Materials and Methods

Participants

Twenty-five native British English-speaking 2.5-year-olds participated in the experiment. Of these, five were excluded because of equipment error (1) and inattention to the screen during the experiment (4; see Section Data Pre-Processing). As a result, 20 toddlers were included in the final set for analysis (Female = 10, mean age = 29 months, range = 25–33 months). Twenty adult participants (native British English-speakers, Female = 13, mean age = 21 years) participated as a control group. No participants reported hearing or vision problems. Ethical approval was obtained via the Ethics Committee for the School of Humanities and Social Sciences of the University of Cambridge.

Apparatus and Procedure

All participants were tested in the Psycholinguistics Lab of the Department of Theoretical and Applied Linguistics in Cambridge, UK. Eye-tracker calibration and stimulus presentation were controlled by Tobii Studio 3.2.1.190 software. The data were obtained with a Tobii X120 infrared eye-tracking camera (Tobii Technology AB; binocular infrared light reflection, 120 Hz sampling frequency, accuracy range: 0.5° to 1°, recovery <300 ms). The camera was placed below a 17-inch monitor and calibrated for distance and angle relative to the monitor. The experimental procedure was the same as for Experiment 1.

Materials

Target sentences

Target sentences were created and spliced using the same procedure as in Experiment 1 (Section Audio Stimuli; **Table 1**). Again, we verified the lexicosyntactic completeness of the targets with a web-based experiment of the sentences in written form (N = 14, female = 8, mean age = 29 years old, native British English speakers). All targets were found to be complete or incomplete, as intended, by more than 75% of the participants. Also as before, we conducted a prosodic completeness listening pre-test (Praat; Boersma and Weenink, 2012; N = 12, female = 10, mean age = 24 years, native British English speakers), which showed that only two (both prosodically complete) target sentences had ambiguous prosody. Their contours were judged as non-interrogative instead of interrogative in more that 25% of the judgments.

Conversation and video construction

Conversations in Experiment 2 (English) were not restricted to word-for-word translations of the conversations in Experiment 1 (Dutch) to allow for more freedom in using child-friendly and culturally appropriate topics (Fenson et al., 1993). However, the structure (turns and placement of conditions) and length (30s) of the conversations were identical between the two experiments.

Audio recordings were obtained using the same procedure from Experiment 1, but with two female native Southern British English speakers (the local dialect in the testing region).

The same pairs of puppets were used from Experiment 1. As before, we created puppet video recordings to match the final audio stimuli. All video recordings were edited for speech alignment and sound quality in Adobe Premiere Elements video editing software. If perfect alignment of sound and movement could not be achieved, the audio signal always preceded the video movement, so that movement of the mouths could not be used as an anticipatory cue for turn transition. This criterion was added to avoid an effect of visual cues to turn transition on participants' looking behavior.

Data Pre-Processing and Analysis

The same criteria and algorithms were used as in Experiment 1 for participant exclusion, anticipatory gaze switche identification, and random-baseline correction of switching values (Section Data Pre-Processing; **Figure 2**). In total, four toddlers (16%) were completely excluded from the analyses for inattention to the screen. No adults were completely excluded. Of the remaining participants (20 toddlers, 20 adults), 27 trials (2.3%) were excluded (25 for the toddlers and two for the adults). The final data set contained gaze data for 1144 trials.

Results

Pre-Analysis: Random Anticipatory Looking

Participants switched less than would be expected by chance in the *fully incomplete* (-SYN -PROS) condition, at chance level for the *incomplete syntax* (−SYN +PROS) condition, and above chance for both the *incomplete prosody* (−SYN +PROS) and *fully complete* (+SYN +PROS) conditions (**Figure 5**). This pattern again suggests that participants use both lexicosyntactic and prosodic cues for turn-projection. As in Experiment 1, when both cues were incomplete, participants were least likely to expect a speaker change, whereas when both cues were complete, they were the most likely to expect one. When the cues were pitted against each other, listeners weighed lexicosyntactic over prosodic cues.

Lexicosyntactic and Prosodic Cues

A model similar to that fitted in Experiment 1 assessed the effects of linguistic cues and age on participants' baseline-corrected anticipatory gaze switches (1144 observations, N = 40;

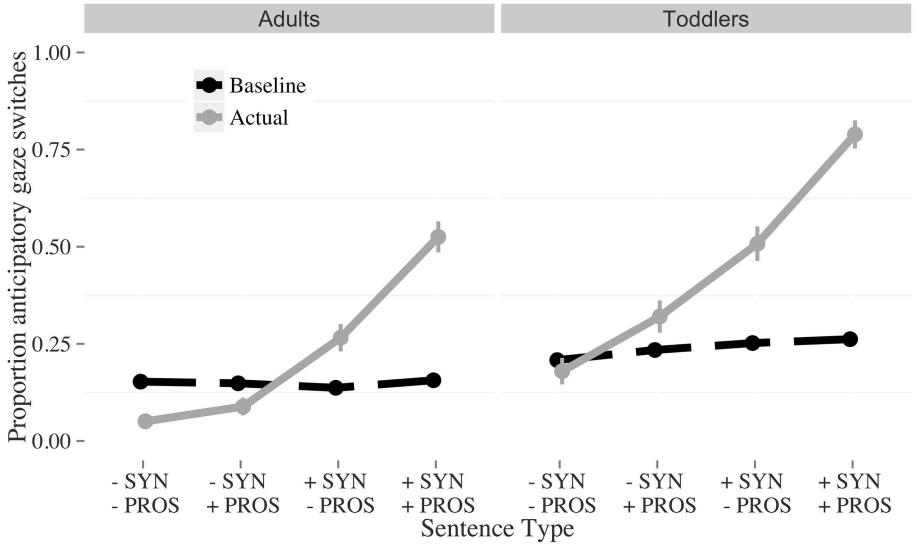


FIGURE 5 | Proportion of baseline (dashed) and actual anticipatory (solid) gaze switches to the answerer, by condition and age. The conditions were: *Fully incomplete* (−SYN −PROS); *Incomplete syntax* (−SYN +PROS); *Incomplete prosody* (+SYN −PROS); *Fully complete* (+SYN +PROS). The vertical bars indicate the standard error of the mean.

TABLE 7 | Outcomes from main linear mixed effects model for both subject groups (English toddlers and adults; Number of observations: 1144, *N* = 40).

Predictor	Contrast coding	β	*t (z)*	*p*
Intercept		0.15	7.88	
Syntactic completeness	Incomplete (−1) Complete (1)	0.17	10.55	<0.001
Prosodic completeness	Incomplete (−1) Complete (1)	0.084	5.12	<0.001
Age	Toddler (-1) Adult (1)	−0.062	−4.37	<0.001
Syntactic completeness × Prosodic completeness		0.044	2.71	<0.05
Syntactic completeness × Age		−0.0084	−0.74	
Prosodic completeness × Age		−0.013	−1.34	

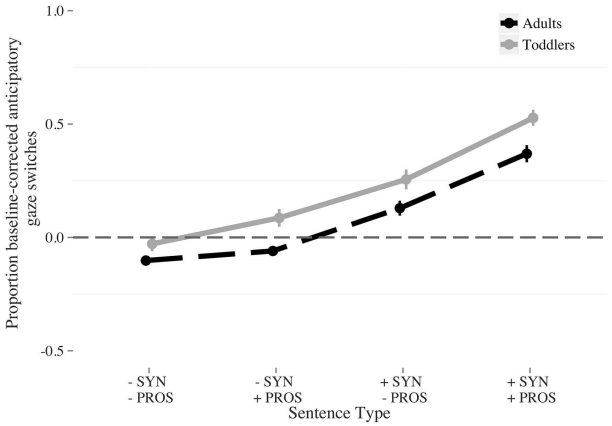


FIGURE 6 | Proportion baseline-corrected anticipatory switches to the answerer by condition and age (Adults: dashed; Toddlers: solid). The conditions were: *Fully incomplete* (−SYN −PROS); *Incomplete syntax* (−SYN +PROS); *Incomplete prosody* (+SYN −PROS); *Fully complete* (+SYN +PROS). The vertical bars indicate the standard error of the mean.

Table 7). The dependent variable was participants' baseline-corrected anticipatory switches. Predictor variables included *syntactic completeness* (incomplete vs. complete), *prosodic completeness* (incomplete vs. complete) and *age* (toddler vs. adult). The intercept was allowed to vary by subject and item, and the predictor variables were contrast-coded (−1, 1).

Again we found that the amount of linguistic information consistent with turn completion affected participants' anticipatory switching (**Figure 6**). Model coefficients show four significant effects (**Figure 6**; **Table 7**). First, the proportion of anticipatory gaze switches was larger for the lexicosyntactically complete vs. lexicosyntactically incomplete targets ($\beta = 0.17$, $z = 10.55$, $p < 0.001$). Second, more anticipatory gaze switches were made for complete prosodic contours vs. incomplete prosodic contours ($\beta = 0.084$, $z = 5.12$, $p < 0.001$). Third, toddlers made more anticipatory gaze switches overall than adults ($\beta = -0.062$, $z = -4.37$, $p < 0.001$). Fourth, there was an interaction between syntactic completeness and prosodic completeness ($\beta = 0.044$, $z = 2.71$, $p < 0.05$). No other coefficients reached significance.

We fit separate *post-hoc* models of the lexicosyntactically complete target sentences (*fully complete* and *incomplete prosody)* and the lexicosyntactically incomplete target sentences (*fully incomplete* and *incomplete syntax*) to explain the interaction between syntactic completeness and prosodic completeness. A model restricted to syntactically complete target sentences (572 observations, $N = 40$; **Table 8**) showed a significant effect of *prosodic*

TABLE 8 | Results from the linear mixed effect models for toddlers and adults together by lexicosyntactic condition (complete or incomplete, number of observations for each condition: 572, *N* = 40).

Lexicosyntactic condition	Complete			Incomplete		
Predictor	***β***	***t(z)***	***p***	***β***	***t(z)***	***p***
Intercept	0.32	11.07		−0.027	−1.30	
Prosodic completeness	0.13	4.92	<0.001	0.040	1.99	0.05
Age	−0.071	−3.22	<0.01	−0.054	−3.61	<0.001
Age × Prosodic completeness	−0.008	−0.44		0.018	−1.33	

completeness, with more anticipatory gaze switches for prosodically complete than prosodically incomplete contours ($\beta = 0.13$, $z = 4.92$, $p < 0.001$). In comparison, a model restricted to syntactically incomplete targets (572 observations, $N = 40$; **Table 8**), only showed a marginal effect of prosodic completeness ($\beta = 0.040$, $z = 1.99$, $p = 0.05$). These *post-hoc* analyses reveal that English listeners' use of prosodic cues depends on whether the utterances are syntactically complete; when utterances were lexicosyntactically incomplete, the effect of prosody was only marginally significant.

TABLE 9 | Outcomes from the linear mixed effects model for the two partially complete conditions for both subject groups (incomplete prosody, incomplete syntax, Number of observations: 572, *N* = 40).

Predictor	Contrast coding	*β*	*t (z)*	*p*
Intercept		0.10	3.63	
Condition	Incomplete syntax (−1) Incomplete prosody (1)	0.089	3.29	<0.01
Age	Toddler (−1) Adult (1)	−0.067	−3.67	<0.001
Condition × Age		0.0045	0.27	

Relative Weight of Lexicosyntactic and Prosodic Cues

Similar to Experiment 1, we then fit a second model restricted to the two partially complete conditions (*incomplete syntax* and *incomplete prosody*) to test the relative weight of lexicosyntactic and prosodic cues (484 observations, $N = 40$; **Table 9**). The predictor variables were *condition* (incomplete syntax vs. incomplete prosody) and *age* (toddler vs. adult).

The model showed that participants weighed lexicosyntactic cues over prosodic cues; they made more anticipatory gaze switches when they only had complete syntax (*incomplete prosody*) compared to when they only had complete prosody (*incomplete syntax*; $\beta = 0.89$; $z = 3.29$, $p < 0.01$).

Speaker Change or Speaker Continuation

We fit an additional model to the baseline-corrected switches in lexicosyntactically complete target utterances (572 observations, $N = 40$; **Table 10**) in order to check whether more anticipatory gaze switches were made when there was a change in speakership compared to when there was no change in speakership. The model suggested no effect of speaker change/continuation ($\beta = 0.039$, $z = 1.53$, $p =$ n.s) Therefore, the effect of lexicosyntax in Experiment 2 cannot be attributed to early visual cues of speaker change (as was possible in Experiment 1).

Finally, we fit two models to data from (a) the first two trials (292 observations, $N = 38$) and (b) the last two trials (252 observations, $N = 33$) of the experiment to test whether children learned to not switch after lexicosyntactically incomplete utterances during the course of the experiment. As in Experiment 1, the *post-hoc* tests showed significant main effects of lexicosyntactic completeness in both the first two trials ($\beta = 0.174$, $z = 4.574$, $p < 0.0001$) and the last two trials ($\beta = 0.183$, $z = 7.252$, $p < 0.0001$, **Table 11**). It is therefore unlikely that our findings for lexicosyntactic completeness were driven by a learned statistical bias.

Discussion

The second experiment showed a very similar pattern of findings to Experiment 1: Toddlers and adults used both lexicosyntactic and prosodic cues for turn projection. Also, both English toddlers and adults weighed lexicosyntactic cues over prosodic cues when the two were pitted against each other.

One difference in the results from Experiments 1 and 2 is that, in Experiment 2 (English), toddlers made more anticipatory gaze switches than adults. This effect of children switching more often than adults has been previously observed in studies with a similar design (Casillas and Frank, 2012), having been explained as an effect of the videos being easy to follow. The explanation is that adults find the videos easy to comprehend and therefore track the turn structure less closely with their eye movements. This explanation fits with our findings in that the presence of ambiguous prosodic contours in Experiment 1 may have made the task more difficult for Dutch adults, leading them to track the conversations more closely than adults did in the clearer, easier contours in the English stimuli.

A second difference between Experiments 1 and 2 is that English listeners used prosodic cues when utterances were lexicosyntactically complete, but not when they were incomplete. This effect was not observed for the Dutch listeners, but it is consistent with prior experimental work on English (Wichmann and Caspers, 2001).

General Discussion

In two experiments, we investigated toddlers' and adults' use of lexicosyntactic and prosodic cues in making predictions

TABLE 10 | Results from the linear mixed effect models for English toddlers and adults together in the lexicosyntactically complete conditions, including the predictor variable speaker change (Number of observations: 572, *N* = 40).

Predictor	Contrast coding	*β*	*t (z)*	*p*
Intercept		0.32	11.23	
Prosodic completeness	Incomplete (−1) Complete (1)	0.13	4.99	<0.001
Age	Toddler (−1) Adult (1)	−0.071	−3.21	<0.01
Speaker change	No (−1) Yes (1)	0.039	1.53	
Prosodic completeness × Age		−0.0076	−0.41	
Prosodic completeness × Speaker change		0.018	−0.70	
Age × Speaker changes		−0.018	−1.01	
Prosodic completeness × Age × Speaker change		−0.014	−0.74	

TABLE 11 | Outcomes from the main linear mixed effects model for (A): first two trials of the experiment and (B): last two trials of the experiment (English toddlers and adults).

	English					
	(A). First two trials 292 observations, *N* = 38			(B). Last two trials 252 observations, *N* = 38		
	β	*t (z)*	*p*	*β*	*t (z)*	*p*
Intercept	0.134	0.0381		0.106	3.973	
Syntactic completeness	0.174	4.574	<0.0001	0.183	7.252	<0.0001
Prosodic completeness	0.0833	2.187	<0.05	0.068	3.827	<0.001
Age	−0.0728	−3.233	<0.01	−0.0351	−1/394	
Syntactic completeness × Prosodic completeness	0.0304	0.797		0.0486	1.920	
Syntactic completeness × Age	0.00680	0.302		−0.0176	0.742	
Prosodic completeness × Age	−0.0114	−0.505		−0.0146	−0.619	
Syntactic completeness × Prosodic completeness × Age	0.0167	0.755		−0.0146	−0.619	

about upcoming turn structure. The experiments were conducted in two languages, Dutch and British English, to test whether the findings were based on language-specific cues for turn prediction. Adults and toddlers in both languages used both lexicosyntactic and prosodic cues in their anticipation of upcoming speaker changes. Participants made the most anticipatory gaze switches when both cues were complete and interrogative (*fully complete*). Participants also anticipated upcoming speaker changes when the lexicosyntactic cue alone was complete and interrogative (*incomplete prosody*). Importantly, complete lexicosyntax alone was not equivalent to the combined effect of complete lexicosyntax and prosody; participants showed a benefit for prosody in that the *fully complete* targets elicited more anticipatory gaze switches than the targets with complete lexicosyntax alone (*incomplete prosody*).

When only the prosodic cue was complete and interrogative (*incomplete syntax*), participants' anticipatory gaze switches did not differ from chance. Participants made the fewest anticipatory gaze switches when both cues were incomplete and declarative (*fully incomplete*), making fewer gaze switches than would be expected by chance. This last finding is the first to demonstrate that toddlers know when *not* to switch; they keep their eyes on the current speaker more often when lexicosyntactic and prosodic cues both signal an incomplete turn.

Our general finding, that listeners use both lexicosyntactic and prosodic cues for turn-projection (but weigh lexicosyntactic information above prosody overall), is compatible with previous findings showing an advantage for *combined* lexical and prosodic cues over lexical cues alone (Duncan, 1972; Ford and Thompson, 1996; Casillas and Frank, 2012, 2013).

Lexicosyntactic vs. Prosodic Cues

We tested the relative weight of lexicosyntactic and prosodic cues by pitting them directly against each other in two conditions (*incomplete syntax* and *incomplete prosody*). Adults were expected to privilege lexicosyntactic information above all (de Ruiter et al., 2006; Magyari and de Ruiter, 2012), while toddlers were expected to privilege prosodic cues instead (Gleitman and Wanner, 1982; Morgan and Demuth, 1996; Jusczyk, 1997; Christophe et al., 2008; Casillas and Frank, 2013). Contrary to our expectations, adults and toddlers did not differ in their relative cue weights; both showed a privilege for lexicosyntactic over prosodic cues in their predictions. There are at least four reasons why this finding could have arisen, three derive from the design of our study and one from the use of prosody for other functions.

Participant's expectations about upcoming turn structure were maximally contrasted for our *fully incomplete* (no speaker switch expected) and *fully complete* (speaker switch expected) conditions. Targets in the *fully incomplete* condition were always

declaratives whereas targets in the *fully complete* condition were interrogatives. Interrogatives automatically cue a speaker switch whereas declaratives don't. As a result, it is important to keep in mind that (by design) the stimuli confounded completeness with interrogative status: both the lexicosyntactic and the prosodic cues to completeness created interrogative utterances, whereas the cues to incompleteness created declarative utterances. Previous work suggests that infants are already sensitive to lexicosyntactic and prosodic cues to questionhood by age two, and that they treat interrogatives differently from declaratives (Lexicosyntax: Shi et al., 2006; Casillas and Frank, 2012, 2013; Geffen and Mintz, 2014; Prosody: Soderstrom et al., 2011; Combined cues: Casillas and Frank, 2012, 2013; Geffen and Mintz, 2012). This pattern continues through adulthood; adults in conversation also give special attention to questions (or other acts eliciting a response; Stivers and Rossano, 2010). In our study, the lexicosyntactic cues to questionhood (subject-auxiliary inversion and *do*-insertion in English) appeared earlier in the utterance than the prosodic cue to questionhood (final high rise). Therefore, it is possible that children weigh lexicosyntactic cues over prosodic cues simply because the lexicosyntactic cues appear earlier than the prosodic ones in the utterances—not because they find lexicosyntactic cues more informative or more important overall. Because our lexicosyntactically complete targets were always formatted as questions (whereas our lexicosyntactically incomplete targets were always formatted as declaratives), the main effect of lexicosyntax could therefore have been driven by a higher response pressure for questions vs. declaratives, instead of for lexicosyntactic complete vs. incompleteness. The current results leave *which* lexicosyntactic cues toddlers used for prediction—completeness, interrogativity, or a combination of the two—as an open question for future research.

As they stand, the current results add to the evidence that toddlers not only distinguish between interrogative and declarative word order (Geffen and Mintz, 2014), but that they are also sensitive to the difference in function between declarative and interrogative utterances. As seen in similar work (Casillas and Frank, 2012, 2013) toddlers made more anticipatory gaze switches after interrogatives compared to declaratives, suggesting that they expect the addressee to reply when a question is (lexicosyntactically) introduced.

A second explanation for toddlers' use of lexicosyntactic cues over prosodic ones is that the lexicosyntactic cues to turn completeness were more consistent in their interpretation (and therefore more reliable) compared to the prosodic cues to turn completeness. Although we took care to select prosodic cues that are relatively consistent and prototypical in signaling a speaker switch (high rising terminal contours to signal interrogativity), rising pitch at prosodic boundaries can, in principle, signal multiple different meanings. There is no one-to-one mapping between intonational contours and their pragmatic function in conversational contexts. Thus, the form-function mappings for prosodic cues may have been less straightforward compared to the mappings for lexicosyntactic cues.

A third, related, explanation derives from a difference in the pragmatic felicity of the two partially complete conditions (*incomplete prosody* and *incomplete syntax*): the *incomplete prosody* condition is less marked than the *incomplete syntax* condition. In natural conversation, it is common for lexicosyntactically complete phases to lack prosodic boundaries (e.g., when the syntactic phrase optionally continues beyond the first possible completion point). But questioning contours rarely occur when lexicosyntax is incomplete, unless they are specifically conditioned by contexts where (a) the addressee is making a repair (Did you mean to say, *"That's a very high"*?) or (b) the speaker is trying to elicit a sentence completion from the addressee, as parents often do with young children during word-elicitation games (e.g., *"A pig says 'oink' and a cow says?"*) As a consequence, it might have been more difficult to understand the *incomplete syntax* target sentences compared to the *incomplete prosody* sentences, thereby explaining the fewer anticipatory gaze switches in the *incomplete syntax* condition without any reference to cue dominance.

A fourth explanation is that prosodic and lexicosyntactic cues are used differently to signal linguistic function, either from the point of view of the *type* of linguistic information being conveyed, or the *extent* of its predictive domain in conversational interaction. Although many studies have shown that children are capable of perceptually distinguishing the types of intonational contours used in the current study (even before the acquisition of segmental and syntactic structure; Snow and Balog, 2002), it is unclear how much of the prosodic system can be acquired before children also master other aspects of the linguistic system. The acquisition of an intonational system involves much more than the ability to produce and discriminate rising and falling pitch movements. Children must also be able to map pitch contours to functional meanings. This involves learning the language's inventory of phonologically distinct intonational contours (e.g., rising, falling, rising-falling, etc…), figuring out what their linguistic and paralinguistic functions are (e.g., rising for interrogativity, but also continuation, etc…), determining how they are realized within utterances (e.g., throughout a phrase, or only in the accented syllable), and finding out what determines variation in their phonetic implementation (e.g., interactions between perceived pitch and fundamental frequency during vowel production). These aspects depend, to a large extent, on other components of the language, namely: metrics, segmental structure, morphosyntax, semantics, information structure, and pragmatics. Therefore, the full acquisition of the prosodic system must be closely intertwined with the development of these other components (cf. Snow, 1994; Oller, 2000). Without these other components, children's predictive prosodic processing is likely to be limited.

Although children get an early start in acquiring prosodic knowledge (compared to lexicosyntactic knowledge), current evidence supports the idea that the acquisition of a full-fledged prosodic system takes many years. While certain aspects of intonational function are acquired in early infancy (e.g., speech act discrimination: Galligan, 1987; Marcos, 1987; Konopczynski, 1995; Prieto and Vanrell, 2007), others remaining elusive even for teenagers (e.g., some implications of nucleus placement and intonation grouping; Cruttenden, 1985). Intonational development has been found to correlate with grammatical

development (e.g., Snow, 1994) and vocabulary size (Chen and Fikkert, 2007). Moreover, recent evidence suggests that children only process prosodic information as intonational phrases once they have acquired a certain amount of syntactic knowledge (phrasal structure; Männel and Friederici, 2010). But it may also be the case that emerging intonation is largely independent of grammatical development, at least for some children (Prieto and Vanrell, 2007).

This explanation extends to the possible predictive value of the prosodic information in our stimuli. The general finding that children's sensitivity to prosodic cues precedes their sensitivity to lexicosyntactic cues has primarily been attested in experimental tasks that tap into more localized functions of prosodic cues, and tend to focus on processing that happens below the level of the utterance (e.g., word segmentation; Nazzi et al., 1998; Grossmann et al., 2005; Christophe et al., 2008). Compared to utterance comprehension in conversational interaction, these experimental tasks operate at a different level of linguistic structure and therefore are likely to utilize somewhat different speech processing mechanisms. Prosodic comprehension in conversational contexts may be substantially different than in experimental contexts, since it is used and understood with interactive goals in mind.

Relatedly, not all prosodic cues are equally useful for predicting upcoming linguistic structure. Prosodic information in isolated linguistic forms, such as a pause or a change in pitch contour, signals a concurrent event (e.g., a syllable with a high pitch as being stressed). In contrast, prosodic information in conversation can also be used to signal *upcoming* events. More specifically, it can be used to predict upcoming prosodic phrase boundaries that can help, in turn, to pick out the intended speech act (e.g., questions vs. non-questions) and to anticipate upcoming turn structure. The use of prosodic information to make predictions in conversation requires that the listener both recognize prosodic phrase boundaries *and* map prosodic contours onto the multitude of possible pragmatic meanings. Evidently, the required linguistic knowledge that underpins the predictive use and interpretation of prosodic information becomes available to children eventually. We therefore suggest that at least some lexicosyntactic information is necessary to put prosodic information to full use in predicting upcoming turn structure during conversation (see also Männel and Friederici, 2010).

This fourth explanation also helps us to interpret the mixed evidence in prior studies about the use of prosodic information for predictive processing (Casillas and Frank, 2012, 2013; Keitel et al., 2013; Keitel and Daum, 2015). Casillas and Frank (2013) found an early, more global role of prosody in turn prediction for 1- and 2-year-olds. In their study, children's predictions only substantially improved with age for utterances with lexically-realized question markers. But, importantly, children still made the most anticipatory gaze switches when *both* prosodic and lexicosyntactic cues were available, suggesting that prosodic knowledge works together with lexicosyntactic information in predicting upcoming turn structure. In Keitel et al.s' (2013) study, only 36-month-olds were able to anticipate upcoming speaker changes and, when they did, they anticipated speaker changes better when intonation was available. Their finding is consistent with the idea that 36-month-olds use both lexicosyntactic and intonational information to predict upcoming speaker changes: 36-month-old children have acquired a substantial amount of lexicosyntactic knowledge that they can use to parse and comprehend intonation, thereby helping them to predict upcoming speaker changes. Adult controls in the same experiment also anticipated upcoming speaker changes, even without the benefit of intonation. But, because other prosodic cues were still present in the pitch-flattened stimuli, the adults in that experiment could have used alternative sources of prosodic information (final lengthening, stress and duration) to make predictions based on prosodic structure, even without intonational contours.

Turn-Projection in a More Natural Context

One of the goals of the study was to investigate the relative weight of lexicosyntactic and prosodic cues in full-signal speech. Prior studies have primarily used phonetic manipulation to remove lexicosyntactic (low-pass filter) and prosodic information (pitch- and duration-resynthesis Casillas and Frank, 2013; Keitel et al., 2013; Keitel and Daum, 2015). One other study used more natural speech materials to control for the presence of lexical cues, but did not control for prosody (Casillas and Frank, 2012). The current study is then the first to test the relative weight of lexicosyntactic and prosodic cues to turn transition in unfiltered, unsynthesized, and thus acoustically full, speech. The current results show that our splicing method is sufficient for investigating the use of lexicosyntactic and prosodic cues on turn prediction in both adults and toddlers.

Though we used full-signal speech, we did not aim for completely realistic stimuli. Instead, by using full-signal speech (like the speech in children's natural environment) we aimed for a balance of experimental control and increased ecological validity. Future studies could further improve the naturalness of the stimuli by making all recording stimuli in spontaneous interactive contexts, instead of pre-scripting the utterances. Read speech differs from natural speech in its prosodic properties in that it has a lower articulation rate, different pause structure (Barik, 1977), and wider pitch range (Eskénazi, 1992) than spontaneous speech. These properties are, in fact, shared with characteristics of infant-directed speech (IDS), the register that used in the present study (Fernald and Simon, 1984; Fernald et al., 1989). However, other prosodic characteristics of read speech are not common in IDS, such as fewer hesitations and fewer rising movements (Levin et al., 1982).

In sum, we showed that the relative weight of linguistic cues in toddler and adult turn projection can be investigated with relatively natural-sounding scripted conversations. Using this technique, we showed that adults and toddlers use both lexicosyntactic and prosodic cues for turn projection, but that lexicosyntactic cues are weighed over prosodic cues when the two are pitted against each other. The results present a challenge for future work to tease apart *which* lexicosyntactic cues children attend to in making their predictions, and how their use of different cues changes throughout development.

Acknowledgments

This research was supported by the European Research Council (Advanced Grant INTERACT #269484 to Stephen C. Levinson), *the Fundatie van de Vrijvrouwe van Renswoude* (to I. Lammertink), and the Radboud University Honours Academy (*Beyond the Frontiers*, to I. Lammertink). We extend our gratitude to those who helped recruit Dutch and English toddlers. We also thank the students that helped voice and videotape the materials, and thank Elma Hilbrink and Stephen C. Levinson for their feedback on earlier versions of this paper.

References

Bakker, M., Kochukhova, O., and von Hofsten, C. (2011). Development of social perception: a conversation study of 6-, 12- and 36-month-old children. *Infant Behav. Dev.* 34, 363–370. doi: 10.1016/j.infbeh.2011.03.001

Barik, H. (1977). Cross-linguistic study of temporal characteristics of different types of speech materials. *Lang. Speech* 20, 116–126.

Bates, D., Maechler, M., and Bolker, B. (2012). *lme4: Linear Mixed-Effects Models Using S4 Classes.* R package version 0.999999-0. Available online at: http://cran.rproject.org/package=lme4

Boersma, P., and Weenink, D. (2012). *Praat: Doing Phonetics by Computer [Computer Software Manual].* Available online at: http://www.praag.org ([Computer Program] Version 5.3.56)

Bögels, S., Magyari, L., and Levinson, S. C. (2014). "Neural correlates of speech preparation in interactive turn-taking: an early start?" in *Poster Presented at Society for Neurobiology of Language 2014* (Amsterdam).

Bruner, J. S. (1975). From communication to language- a psychological perspective. *Cognition* 3, 255–287. doi: 10.1016/0010-0277(74)90012-2

Casillas, M., Bobb, S., and Clark, E. (in press). Turn-taking, timing and planning in early language acquisition. *J. Child Lang.*

Casillas, M., and Frank, M. C. (2012). "Cues to boundary prediction in adults and preschoolers," in *Paper presented at the Proceedings of SemDial 2012: The 16th workshop on Semantics and Pragmatics of Dialogue* (Paris: Universite Paris-Diderot).

Casillas, M., and Frank, M. C. (2013). "The development of predictive processes in children's discourse understanding," in *Proceedings of the 35th Annual Meeting of Cognitive Science Society* (Austin, TX).

Caspers, J. (2001). "Testing the perceptual relevance of syntactic completion and melodic configuration for turn-taking in Dutch," in *Paper presented at the Proceedings of the 7th European Conference on Speech Communication and Technology* (Aalborg).

Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *J. Phon.* 31, 251–276. doi: 10.1016/S0095-4470(03)00007-X

Chen, A., and Fikkert, P. (2007). "Intonation of early two-word utterances in Dutch," in *Proceedings of the International Congress of the Phonetic Sciences XVI,* (Saarbrucken), 315–320.

Christophe, A., Millotte, S., Bernal, S., and Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Lang. Speech* 51(Pt 1–2), 61–75. doi: 10.1177/0023830908510010501

Couper-Kuhlen, E. (2012). "Some truths and untruths about final intonation in conversational questions," in *Questions,* ed J. P. de Ruiter (Cambridge: Cambridge University Press), 123–145.

Cruttenden, A. (1985). Intonation comprehension in 10-years-olds. *J. Child Lang.* 12, 643–661. doi: 10.1017/S030500090000670X

de Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

Dryer, M. S. (2013). "Order of subject and verb," in *The World Atlas of Language Structures Online,* eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology). Available online at: http://wals.info/chapter/82 (Accessed July 4, 2014).

Duncan, S. (1972). Some signals and rules for speaking turns in conversations. *J. Pers. Soc. Psychol.* 23, 283. doi: 10.1037/h0033031

Englert, C. (2010). Questions and responses in Dutch conversations. *J. Pragmat.* 42, 2666–2684. doi: 10.1016/j.pragma.2010.04.005

Ervin-Tripp, S. (1979). "Children's verbal turn-taking," in *Developmental Pragmatics,* eds E. Ochs and B. B. Schieffelin (New York, NY: Academic Press), 391–414.

Eskénazi, M. (1992). "Changing speech styles: strategies in read speech and casual and careful spontaneous speech," *in Second International Conference on Spoken Language Processing* (Banff, AB).

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manuals.* Baltimore, MD: Paul H. Brokes Publishing.

Fernald, A., and Simon, T. (1984). Expanded intonation contours in mother's speech to newborns. *Dev. Psychol.* 27, 104–113. doi: 10.1037/0012-1649.20.1.104

Fernald, A., Swingley, D., and Pinto, J. P. (2001). When half a word is enough: infants can recognize spoken words using partial phonetic information. *Child Dev.* 72, 1003–1015. doi: 10.1111/1467-8624.00331

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang.* 16, 477–501. doi: 10.1017/S0305000900010679

Ford, C. E., and Thompson, S. A. (1996). "Interactional units in conversation: syntactic, intonational and pragmatic resources for the management of turns," in *Interaction and Grammar,* eds R. Ochs, E. A. Schegloff, and S. Thompson (Cambridge: Cambridge University Press), 134–184.

Galligan, R. (1987). Intonation with single words: purposive an grammatical use. *J. Child Lang.* 14, 1–21. doi: 10.1017/S0305000900012708

Garvey, C. (1984). *Children's Talk.* Cambridge, MA: Harvard University Press.

Garvey, C., and Berninger, G. (1981). Timing and turn-taking in children's' conversations. *Discourse Process.* 4, 27–57. doi: 10.1080/01638538109544505

Geffen, S., and Mintz, T. H. (2012). "Seven-months-olds' discrimination of statements and questions," in *Paper Presented at the Proceedings of the 36th Annual Boston University Conference on Language Development* (Boston,CA).

Geffen, S., and Mintz, T. H. (2014). Can you believe it? 12-Month-olds use word order to distinguish between declaratives and polar interrogatives. *Lang. Learn. Dev.* 1, 15. doi: 10.1080/15475441.2014.951595

Gervain, J., and Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nat. Commun.* 4:1490. doi: 10.1038/ncomms2430

Ginsburg, G. P., and Kilbourne, B. K. (1988). Emergence of vocal alternation in mother-infant interchanges. *J. Child Lang.* 15, 221–235. doi: 10.1017/S0305000900012344

Gleitman, L. R., and Wanner, E. (1982). "The state of the state of the art," in *Language Acquisition,* eds E. Wanner and L. R. Gleitman (Cambridge: Cambridge University Press), 3–48.

Grabe, E., and Post, B. (2004). "Intonational variation in the British Isles," in *Corpus Linguistics: Readings in a Widening Discipline,* eds G. Sampson and D. McCarthy (London; New York: Continuum International), 474–481.

Grossmann, T., Striano, T., and Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *Neuroreport* 16, 1825–1828. doi: 10.1097/01.wnr.0000185964.34336.b1

Gunlogson, C. (2001). "Rising declarative questions," in *Paper Presented at the Proceedings of the Formal Pragmatics Conference* (Berlin).

Haan, J. (2002). *Speaking of Questions: an Exploration of Dutch Question Intonation.* Dissertation, University of Utrecht, Utrecht.

Hirvenkari, L., Ruusuvuori, J., Saarinen, V.-M., Kivioja, M., and Peräkylä, A. (2013). Influence of turn-taking in a two-person conversation on the gaze of a viewer. *PLoS ONE* 8:e71569. doi: 10.1371/journal.pone.0071569

Höhle, B., and Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Dev. Sci.* 6, 122–127. doi: 10.1111/1467-7687.00261

Holler, J., and Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction:optimizing recipiency. *Front. Psychol.* 6:98. doi: 10.3389/fpsyg.2015.00098

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT press.

Keitel, A., and Daum, M. M. (2015). The use of intonation for turn anticipation in observed conversations without visual signals as source of information. *Front. Psychol. Lang. Sci.* 6:108. doi: 10.3389/fpsyg.2015.00108

Keitel, A., Prinz, W., Friederici, A. D., von Hofsten, C., and Daum, M. M. (2013). Perception of conversations: the importance of semantics and intonation in children's development. *J. Exp. Child Psychol.* 116, 264–277. doi: 10.1016/j.jecp.2013.06.005

Konopczynski, G. (1995). "A developmental model of acquisition of rhythmic patterns: results from a cross-linguistic study," in *Proceedings of the XIIIth International Congress of Phonetic Sciences 4* (Stockholm).

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT press.

Levin, H., Schaffer, C. A., and Snow, C. (1982). The prosodic and paralinguistic features of reading and telling stories. *Lang. Speech* 25, 43–54.

Levinson, S. C. (2013). "Action formation and ascription," in *The Handbook of Conversation Analysis*, eds T. Stivers and J. Sidnell (Malden, MA: Wiley-Blackwell), 103–130.

Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

Männel, C., and Friederici, A. D. (2010). Intonational phrase structure processing at differentstages of syntax acquisition: ERP studies in 2-,3-, and 6-year-old children. *Dev. Sci.* 14, 786–798. doi: 10.1111/j.1467-7687.2010.01025.x

Marcos, H. (1987). Communicative functions of pitch range and pitch direction in infants. *J. Child Lang.* 14, 255–268. doi: 10.1017/S0305000900012915

Morgan, J., and Demuth, K. (1996). "Signal to syntax: an overview," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition,* eds J. Morgan and K. Demuth (Mahwah, NJ: Lawrence Erlbauw Associates), 1–22.

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 756–766. doi: 10.1037/0096-1523.24.3.756

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum and Associates.

Pisoni, D. B. (1981). Speeded classification of natural and synthetic speech in a lexical decisiontask. *J. Acoust. Soc. Am.* 70, S98. doi: 10.1121/1.2019150

Prieto, P., and Vanrell, M. M. (2007). "Early intonational development in Catalan," in *Proceedings of the International Congress of the Phonetic Sciences XVI,* eds J. Trouvain and W. J. Barry (Dudweiler: Pirrot GmbH).

R Development Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org

Rossano, F., Brown, P., and Levinson, S. C. (2009). "Gaze, questioning and culture," in *Conversation Analysis: Comparative Perspectives,* ed J. Sidnell (Cambridge University Press: Cambridge), 187–249. doi: 10.1017/CBO9781631635670.008

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language* 5, 696–735. doi: 10.1353/lan.1974.0010

Shi, R., Werker, J. F., and Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy* 10, 187–198. doi: 10.1207/s15327078in1002_5

Shi, R., Werker, J. F., and Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* 72, B11–B21. doi: 10.1016/S0010-0277(99)00047-5

Snow, C. E. (1977). The development of conversation between mothers and babies. *J. Child Lang.* 4, 1–22. doi: 10.1017/S0305000900000453

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *J. Speech Hear. Res.* 37, 831–840. doi: 10.1044/jshr.3704.831

Snow, D., and Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua* 112, 1025–1058. doi: 10.1016/S0024-3841(02)00060-8

Soderstrom, M., Ko, E., and Nevzorova, U. (2011). It's a question? Infants attend differently to yes/no questions and declaratives. *Infant Behav. Dev.* 34, 107–110. doi: 10.1016/j.infbeh.2010.10.003

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Helnemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Stivers, T., and Rossano, F. (2010). Mobilizing response. *Res. Lang. Soc. Interact.* 43, 3–31. doi: 10.1080/08351810903471258

Thorgrímsson, G., Fawcett, C., and Liszkowski, U. (2015). 1- and 2-year-olds' expectations about third-part communicative actions. *Infant Behav. Dev.* 39, 53–66. doi: 10.1016/j.infbeh.2015.02.002

Tice, M., and Henetz, T. (2011). "Turn-boundary projection: looking ahead," in *Proceedings of the 33nd Annual Meeting of the Cognitive Science Society* (Boston, MA).

Wichmann, A., and Caspers, J. (2001). "Melodic cues to turn-taking in English: evidence from perception," in *Paper presented at the Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue* (Aalborg). doi: 10.3115/1118078.1118103

Zink, I., and Lejaegere, M. (2003). *Aanpassingen en hernormering van de MacArthur CDI's van Fenson et al., 1993.* Acco.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ADVANTAGES OF PUBLISHING IN FRONTIERS

FAST PUBLICATION
Average 90 days from submission to publication

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

RESEARCH NETWORK
Our network increases readership for your article

OPEN ACCESS
Articles are free to read, for greatest visibility

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

GLOBAL SPREAD
Six million monthly page views worldwide

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

IMPACT METRICS
Advanced metrics track your article's impact

SUPPORT
By our Swiss-based editorial team

frontiers

EPFL Innovation Park · Building I · 1015 Lausanne · Switzerland
T +41 21 510 17 00 · info@frontiersin.org · frontiersin.org