TIMING TURNS IN CONVERSATION
A TEMPORAL PREPARATION ACCOUNT

ISBN 978-90-76203-67-6
Cover illustration: Szabó Evu
Printed and bound by Ipskamp Drukkers b.v., Nijmegen

TIMING TURNS IN CONVERSATION
A TEMPORAL PREPARATION ACCOUNT

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 9 juli 2015
om 16.30 uur precies

door
Lilla Magyari
geboren op 30 januari 1981
te Budapest, Hongarije

Promotoren:
Prof. dr. S. C. Levinson
Prof. dr. J.P. de Ruiter (Bielefeld University, Duitsland)

Copromotor:
Dr. M.C.M. Bastiaansen (NHTV Breda)

Manuscriptcommissie:
Prof. dr. P. Hagoort
Prof. dr. M.B. Jansma (Universiteit Maastricht)
Prof. dr. J. van Berkum (Universiteit Utrecht)

The research reported in this thesis was supported by
the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany.

To Clair, Kaoru and Sylvia

ACKNOWLEDGEMENTS

I would like to thank the people who have supported me on this long journey towards the completion of this book. First of all, I want to thank my promotors, Steve Levinson, Jan de Ruiter and Marcel Bastiaansen who guided me in my PhD project through all these years.

It was a great privilege to have the opportunity to start such an exciting research project in the Language & Cognition Group led by Steve Levinson. Steve's insights, comments and encouragement were extremely helpful along the way. He has showed me that science can be difficult and challenging, but also exciting and fun at the same time. I feel also honoured for the time and attention he devoted to my writing in order to complete this book.

I was lucky to start my PhD under JP's supervision. I have learned so much from him about experimental psychology. This knowledge still supports me in my work every day. I also enjoyed his humour very much and it helped me through a lot of difficult moments.

Being able to work with Marcel on my master's thesis and later on my dissertation was an irreplaceable help for me. I have benefitted tremendously from Marcel's experience with electrophysiology. Above all his positive attitude made me more confident in my work.

I would also like thank others who supported me with knowledge, experience, opportunities and scientific inspiration. First of all, I thank Csaba Pléh whose works and talks have always been a source of inspiration for me. I thank him so much for all his continuous support and interest in my work even after I left Hungary.

The doors of the Neurobiology of Language group were always open for me for which I am very grateful. I especially thank Jos van Berkum for the helpful discussions and Peter Hagoort for the opportunity to work as an assistant in the NBL group.

I can not express how much I owe Robert Oostenfeld and the FieldTrip Development Team. Special thanks to Robert for his encouragement and generosity sharing his knowledge (and for patiently explaining things over and over again).

The members of the L&C group were the ones who helped, discussed and commented on my work on a daily bases. I thank Tanya Stivers for introducing me to Conversation Analysis and Nick Enfield, Asifa Majid, Gunter Senft for their attention when I had a question about academic life or about other subjects, Kobin Kendrick, Mark Dingemanse for our scientific discussions and Edith Sjoerdsma for helping me with any kind of problem. I especially thank Sara Bögels for the exciting discussions on turn-taking and EEG.

I am also grateful for the excellent support of the Technical Group at MPI. Special thanks to Alex Dukers and Herbert Baumann for their generous help and patience with my desperation caused by technical difficulties.

The members of the L&C group, their partners and others from the MPI were not only academically important but they were my "family" and friends in Nijmegen. I thank Lila, Chris, Elisabeth, Alejandro, Simeon, Rosa, Jeremy, Kat, Harald, Theresa, Federico, Pamela, Luke, Hilário, Neil, Benedicte, Paula, Marloes, Giusy, Nick, Francisco, Julia, Daniel, Kendall, Ingun, Ewelina, Hedvig, Ruth, Michael and others for their emotional support and for their time to have conversation and beers (and once for performing a theater-play with me!). I also thank my friends outside of MPI, Tim, Janneke, Tarik, Didier, Tilman, Liesbeth and Mirna for helping me distract myself from work with fun. I especially thank Peter Nijland who has been a good friend and also an excellent assistant in running some of my experiments. I was lucky to have the opportunity to dance in the Wednesday Morning Group in the last years while working on my dissertation. Our sessions and friendship inspired me in many unexpected, unimaginable ways related to both my scientific and artistic ideas.

Simeon, Elisabeth, Elliott, Luke, Clair and Tim (and I am afraid the list is not exhaustive) deserve a special thanks because they proofread many parts and versions of this dissertation and corrected my English mistakes.

I would like to express my gratitude to Birgit and Mundi, my paranyphms. I thank Birgit that I could (and I always can) count on her whenever I need help. Mundi's support and friendship also helped me a lot in the last few years. I will never forget our conversations and improvised music sessions at his beautiful place in the Ooipolder.

I feel honoured that many of my Hungarian friends have not lost contact with me, and stayed a stable point in my life through all the years writing my dissertation in the Netherlands. I thank Sáriék, Zümiék, Nóriék, Milánék, Doorsék, Csucsuék, Esztiék, Rékaék, Dzsíék, Bandi, Sanyi, Zsófi, Zsuzsa and Kornél. Especially, I thank Évu who gave such an elegant final shape and cover design to this book.

I also thank my family, Anyu, Laci, Apu, Edit, Gréta, Péter, Glenda, Patrik, Gabi and Dodó for their support. It was extremely important for me to feel their warm welcome anytime I went home and their pride when I was away. I thank my Mom who never gave up trying to contact me even if I forgot to call her back for days, and my sister, Gréta and her partner, Péter who visited me so often in Nijmegen. I also thank my Dad who was an inspiration for me to explore the yet unknown.

I can never thank my boyfriend, Tim, enough. He witnessed and comforted me in many of my desperate moments during the last years of working on this book. Without his support and love I could have hardly finished this dissertation.

Finally, I would like to thank my friends and colleagues, Clair, Kaoru and Sylvia who started their PhD together with me in the L&C group and to whom this book is dedicated. For years, we have shared the up and down moments of the PhD life and supported each other. Our friendship is a "side-effect" of this dissertation, an unexpected, beautiful present. I shall remain ever grateful for it.

CONTENTS

CHAPTER 1

1 GENERAL INTRODUCTION

The dissertation explores some of the cognitive processes underlying our ability to participate in everyday natural conversations. In particular, it examines how speakers time their turns to closely fit with the end of the previous turn.

Although everyday conversations run smoothly and effortlessly most of the time, a closer look at the timing pattern of conversations implies a complex cognitive architecture underlying the timing of conversational turn-taking. The significance of studying timing of turn-taking is twofold. First, the temporal patterning of turn-taking poses a challenge for traditional language processing models because it forces us to consider how language comprehension, speech production and other cognitive processes interact to facilitate conversation—the primary ecological niche for language. And second, it highlights a fundamental human skill for timing, a crucial element in human social interactions that has rarely been examined in the context of language use.

This thesis provides an initial study of the cognitive and neurocognitive processes in the timing of turn-taking, guided by the following hypothesis: Speakers predict when a turn ends by predicting how it ends.

1.1 The issue

Natural conversations are everyday verbal interactions involving two or more speakers who alternate freely in speaking (Sacks, Schegloff & Jefferson, 1974; Levinson, 1983). Interactants in natural conversations switch rapidly between the roles of listener and speaker without any overt external constraints regulating who speaks and for how long. Despite the absence of explicit regulatory constraints, however, the internal structure of conversation and the temporal patterns of turn-taking exhibit an underlying systematic organization (Sacks et al., 1974; Stivers et al., 2009). Regarding the timing pattern, interactants show sensitivity to how fast their partner responds. Delays and overlaps are interactionally consequential. For example, a short silence before a turn can be understood as presaging disagreement (Pomerantz, 1984; Levinson, 1983). When overlap occurs, it is normally resolved by speaker withdrawal such that only one speaker remains (Schegloff, 2000). This suggests that the "default" mode of conversation is aimed at avoiding long gaps and overlaps (Sacks et al., 1974). This has been supported by recent corpus studies which have measured the timing of turn-transitions on a millisecond scale. These studies have found that the duration of turn-transitions (i.e. the time between the end of a turn and the beginning of the next one) is most frequently between 0-200 ms (e.g. Stivers et al., 2009; Heldner & Edlund, 2010). This means that interactants typically switch from listening to speaking in less than one fifth of a second.

Such short turn-transition times are surprising from a cognitive processing point of view. Listeners who are to speak next must accomplish several tasks in this short time window. They must, at the very least, sufficiently comprehend the turn in progress while also preparing for the production of their own upcoming turn. This short time window is even more surprising considering the relatively long latency of the speech production process. Studies using picture naming tasks indicate that about 600 ms of preparation time is needed before a single word begins to be articulated (see Levelt, 1989; Indefrey & Levelt, 2004; Indefrey, 2011)—far more than the 0-200 ms interval observed in natural interactions.

The intriguing psycholinguistic question, then, is how the speech production and comprehension systems enable the regular achievement of smooth and fluent turn-transitions in everyday conversations given the tight timing that is observed. Fast turn-transitions suggest that participants cannot just wait until the other has finished speaking and then start to speak, but must start planning in advance (Levinson, 2013). Planning of a next turn can probably only occur if the listener (next speaker) can anticipate the content of a current turn before it finishes. Research has long suggested that speakers must also predict when a given turn might end, so that their turns can be timed closely to the end of the previous turn (Duncan, 1974; Sacks et al., 1974).

The central issue of the dissertation is the relationship between these two processes: anticipation of a turn's content, and the prediction of when the turn ends. Most of the studies of the dissertation investigate whether prediction of the CONTENT enables not only the preparation of the next turn but also the prediction of WHEN the turn ends.

Up to now, two different accounts of how participants predict turn endings have been proposed: (1) via reaction to turn-yielding cues, and (2) via prediction of the linguistic constructions that constitute the turn. The first account holds that observable cues appear in the speaker's speech or behaviour shortly before turn-endings, and that these signal to listeners that the turn is coming to an end (e.g. Duncan, 1974; Duncan & Fiske, 1977). By contrast, the other account, proposed by Sacks et al. (1974), contends that listeners predict the type of construction that a speaker is going to produce (e.g. a word, phrase, clause, or multi-clausal construction) to estimate when the turn will likely end.

These accounts implicate two different mechanisms supporting precision timing in conversations: one reactive, and one anticipatory. On the reactive turn-yielding account, perceptual cues signal incipient turn-ending to the listener, who reacts by initiating the next turn. In contrast to this approach, the alternative proposal is that grammatical formatting aids anticipation of turn-endings. For example, a turn beginning with *if (whenever, either,* etc.*)* projects a likely two-clause structure downstream.

Using such syntactic projection, the listeners can assume that the duration of this turn will probably take longer than the duration of a single phrase construction. The anticipatory mechanism may not involve such distal projection of course, but are arguably distinct from late turn-final cues. The anticipated lexical information (i.e. anticipation of the words contained in the turn) might also provide information about turn duration in advance.

This thesis will focus on this latter account of turn-end predictions. More precisely, most of the studies in this dissertation examine whether prediction of the (lexical and syntactic) content of turns facilitate precise temporal predictions of turn-ends.

1.2 Earlier research

In order to better understand the cognitive processes underlying the short turn-transitions of everyday conversations, the time-courses of language comprehension and production processes are especially relevant. These processes have traditionally been studied separately and therefore, they are discussed in the two following sections (though see e.g., Hagoort, 2014; Hagoort & Indefrey, 2014, Pickering & Garrod, 2013 for integrative accounts of comprehension and production). Subsequently, the literature on language processing in interaction is reviewed.

1.2.1 Predictive comprehension

Recent developments in cognitive, computational and neuroscience research show that a universal operation of the brain is the continual generation of predictions (Bar, 2009; Friston, 2010). Recent models of language processing also emphasize the incremental and predictive nature of language comprehension (Kutas, DeLong & Smith, 2011; Pickering & Garrod, 2013; Hagoort & Indefrey, 2014, also see Van Berkum, 2013). Using the high time resolution of eye-tracking and EEG, studies have revealed predictions made at different levels of language comprehension. In the visual word paradigm of eye-tracking studies, participants' eye-movements are recorded as they observe a visual scene and listen to sentences that refer to objects in that scene (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). These studies demonstrate that participants look to possible referents in the visual scenes prior to their being mentioned in the sentence (e.g. Kamide, Altmann & Haywood, 2003; Knoeferle, Croecker, Scheepers & Pickering, 2005; Altmann & Kamide, 2007). This implies that participants combine their semantic analysis of the incoming speech with the visual context early on, thereby narrowing the possible trajectories the speech might take and affording predictive understanding. ERP studies have also demonstrated that participants predict upcoming words during sentence processing by manipulating expectations of lexical gender or phonological form (Wicha, Moreno & Kutas, 2004; DeLong, Urbach & Kutas, 2005; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005). In sum, empirical evidence shows that listeners make predictions during language comprehension. Listeners not only predict developments in the situation under discussion, but can also predict specific words (also see Van Berkum, 2013).

Predictions facilitate the speed of language comprehension, and can help to disambiguate noisy input in natural language use (see e.g. Kutas et al., 2011; Pickering & Garrod, 2007). For example, contextually based predictions greatly reduce the number of lexical candidates activated during the processing of an incoming acoustic signal. Therefore, word recognition happens quickly and is completed within a few hundred milliseconds (see Hagoort & Poeppel, 2013). Furthermore, the comprehension system is not only able to facilitate the processing of linguistic information by incremental analysis, but it also processes information which has already been predicted but not yet encountered in the linguistic input (Kutas et al., 2011).

1.2.2 Latencies in production

In the last thirty years, experimental research has also targeted the time-course of different stages of speech production. There is some consensus about four major stages of the speech production system (conceptual preparation, lexical access, phonological processing, and articulation), although the architectural details are debated (Dell, 1986; Levelt, 1989; Caramazza, 1997). Detailed studies of word production have shown that, given a picture to name, it takes from 600 ms (for simple pictures depicting frequent words) to 1200 ms (for infrequent words) to retrieve and code a word in preparation for articulation (Levelt, 1989:222). Naming latencies can vary with variations in the task or the stimuli, however. For example, repetition of the same word, word length, familiarity, word-frequency, priming effects or cognate status also influence naming times (e.g. Jescheniak & Levelt, 1994; Jescheniak, Schriefers & Hantsch, 2003; Strijkers, Costa & Thierry, 2009). More recently, naming latencies have been reported in the range of 470 to 2000 ms (see Indefrey, 2011).

Natural language use, however, is characterized not by the production of single words, but by grammatical constructions of varying length. Accordingly, if speakers must prepare larger units before articulation, we might expect a significant drag on preparation for speech production. In the speech production literature, there is a discussion about the size of the planning units before starting articulation. Their proposals vary from radical incrementality, where only one word is planned at a time (Levelt & Meyer, 2000; Gleitman, January, Nappa & Trueswell, 2007), to the generation of the structural frame before production (Griffin & Bock, 2000; Bock, Eberhard & Cutting, 2004) and the pre-activation of the phonological form of multiple words (Costa & Caramazza, 2002). More recent studies argue that the time-course of formulation is flexible, and speakers might use planning units of different sizes under different circumstances (e.g. Wagner, Jescheniak & Schriefers, 2010; Konopka & Meyer, 2014). Nonetheless, even if speakers need to prepare only the first word of a turn before starting articulation, the time-course of single word production is still relatively slow compared to the tight timing of conversational turns.

The time needed for the conceptual preparation of speech in picture naming studies, however, may differ from the time needed for the conceptual preparation of conversational turns. Meta-analyses of picture-naming studies provide insight into the estimated duration of the different stages of word production. These studies estimate that at least 200 ms are required for conceptual preparation for word production (Indefrey & Levelt, 2004; Indefrey, 2011). Thus, the duration from the end of conceptual preparation to articulation was estimated as at least 400 ms. This duration is still twice as long as the average gap between turns in conversational interaction.

1.2.3 Turn-taking models and facts

Experimental studies on verbal interaction between individuals have typically focused either on interactants' perspective-taking (e.g. Clark & Wilkes-Gibbs, 1986; Horton & Keysar, 1996; Barr, 2008), or on their linguistic coordination by lexical and structural priming (Branigan, Pickering & Cleland, 2000; Pickering & Garrod, 2004, but see Healey, Purver & Howes, 2014 for an alternative claim which questions the role of structural priming in conversation).

Surprisingly, there has been relatively little experimental work on how interactants achieve the fast switch from comprehension to production "mode". However, the time-course of these processes is intriguing for any models of language processing. Single word production studies (1.2.2) show that the speech production process is relatively slow compared to the tight timing of turn-transitions. At present, we do not know how the results of earlier studies might carry over to interactive contexts. For example, the 600 ms minimum for single word production might be reduced by incremental processing or strong contextual effects in a conversational setting—although we can say that the last 400 ms of that interval (i.e. the duration from the end of the conceptual preparation to articulation) is likely to be a hard barrier. Additionally, conversation imposes many additional demands absent from single-word studies that could lead to longer conceptual preparation times (e.g. selecting the appropriate word for a particular addressee or topic). Hence, it is likely that planning for a next turn often precedes the ending of the current turn in natural conversations.

It is not obvious how listeners can start to prepare an answer to a turn which has not yet been fully completed. One possibility is that relatively early on listeners predict the likely trajectory of a current turn in progress. Earlier research (reviewed in 1.2.1) shows that listeners do indeed make predictions about the upcoming speech, and also process predicted-but-not-yet-encountered information. This suggests that preparation of an answer could start as soon as the content of the current turn is sufficiently predictable. Little is known, however, about "how much" comprehension is needed before starting to prepare a response. Further research is necessary to establish how much listeners need to hear to anticipate the probable content of a given turn, and whether they need to predict the actual words and syntactic structures to prepare an appropriate response. For example, Levinson (2013) suggests that action-specific features might appear early in turns and can help to start preparation of an answer. In accordance with this, Gísladóttir and colleagues (2012) showed that speech acts can sometimes be recognized at the first word of utterances.

Another possibility could be that speakers do not need to predict the content of a current turn because they can start their turn with fillers or particles (*well, um, uh,* etc.) that may be used independently of the content of a previous turn. In this case, speakers could start preparing their turn without fully understanding the current turn's content. However, conversation analytic and corpus studies (e.g. Clark & Fox Tree, 2002) have robustly shown that there are no components which are entirely unrelated to the action that a given turn implements. For example, hesitations and particles like *well* in English are likely to appear at the beginning of turns which do not wholly conform to the expectations set by a preceding turn (Pomerantz, 1984; Kendrick & Torreira, 2014). This means that speakers probably design turn-beginnings (i.e. the place where fillers and particles often appear) already with an idea of what sort of action they will implement through their turn.

To summarize, the estimated latency of speech production suggests that planning for a next turn must start before the current turn ends. This is possible if we assume that listeners (the next speakers) can predict the content of the current turn, and therefore, can start to prepare the next turn. Accordingly, experimental evidence indeed shows that listeners make predictions at many levels of the speech comprehension process.

The issue taken up in the next two sections is whether the observed tight timing of turn-transitions is not only explained by an early start of the speech production process but also by the prediction of turn-endings. Some researchers argue that it is unnecessary to assume precise timing of turns in order to explain the observed distribution of the duration of turn-transitions (Heldner & Edlund, 2010). In contrast with this, Sacks et al. (1974) argued that listeners must predict the end of conversational turns in order to enter without much gaps or overlaps. Since then, two schools of thought have arisen related to the processes which enable the prediction of turn-ends. On the one hand, researchers have proposed that listeners use the syntactic frame and the intonational envelope to predict the overall structure of the incoming turn (see 1.2.3.2.). Others characterized possible turn-yielding cues which appear just before turn-ends and are assumed to signal that the speaker will finish the current turn soon (see 1.2.3.1.).

1.2.3.1. Turn-yielding cues

Turn-yielding cues are those perceptual features of behavior that appear towards the end of conversational turns and signal that the current turn is coming to an end. Most of the suggested cues are prosodic, for instance, final syllable lengthening or pitch changes in the last word (e.g. Duncan, 1974; Duncan & Fiske, 1977; Schaffer, 1983; Local, Wells & Sebba, 1985; Local, Kelly & Wells, 1986, Schegloff, 1996). Other research has also proposed non-verbal signals, for example, particular kinds of eye-gaze and gesture (Kendon, 1967; Duncan, 1974).

With regard to non-verbal behaviour, Kendon (1967) found that eye-gaze is systematically coordinated with the timing of speech. For example, if the speaker does not look up at the end of an utterance, there is a longer gap before the reply. More recently, precise measurements of turn-transition durations have revealed that fast transitions are frequent regardless of whether the conversation takes place face-to-face or in a telephone-like situation (Ten Bosch, Oostdijk & De Ruiter, 2005; De Ruiter et al., 2006; Stivers et al., 2009; but already noted in Levinson, 1983:302). Other work by Rossano (2012) also points out that gaze-behaviour is influenced more by the structural organization of social actions in conversations (i.e. by "sequence organization" as understood in Conversation Analysis) than by turn-taking. These findings undermine the role of non-speech cues as a major factor in turn-end predictions, and suggest that the mechanisms related to auditory speech processing should account for how listeners predict turn-ends.

With regard to prosody, studies have typically focused on turn-final intonation contours (e.g. Local, et al., 1985). Some experimental work has studied whether listeners perceive pitch changes as turn-yielding (Beattie, et al., 1982; Schaffer, 1983; Cutler & Pearson, 1986). For example, Beattie and colleagues (1982) examined interviews with Margaret Thatcher. They concluded that she was interrupted often in conversations because she used sharply dropping intonation contours in the middle of turns, where such contours were perceived as turn-ending signals. More recently, De Ruiter and colleagues (2006) attempted to disentangle prosody and lexico-syntactic information in turn-end predictions. They played recordings of turns from natural conversation to participants and asked them to press a button when a turn ended. They modified the recordings so that either the intonation contour or the lexical information (or both) were missing from the recordings. When participants listened to turns without the intonation contour, there was no change in the accuracy of their button-presses as compared to their performance with the original recordings. But when the words were obscured (and intonational contour remained intact), participants' performance was significantly worse. They concluded that intonation was neither sufficient nor necessary for prediction of turn-ends, and that syntactic and lexical information played a major role in timing of turn-taking. However, this study did not take into account the use of non-pitch prosodic information. As already mentioned, the use of word final lengthening has also been suggested to signal turn-ends (Local & Walker, 2012; see also White, 2014), and there are many potential prosodic and articulatory cues besides pitch (see Bögels & Torreira, submitted).

Wilson and Wilson (2005) suggest that precise turn-transitions cannot be solely explained by the listener's perception of turn-yielding cues. They propose that speakers also become mutually entrained on the basis of the current speaker's rate of syllable production. This entrainment leads to an oscillatory cycle of readiness to initiate speech. Wilson & Wilson propose that very short gaps are common in conversations because the oscillatory functions of the current and the next speaker are counter-phased. The next speaker's readiness to initiate speech is at the maximum when the current speaker is mid-syllable. Hence, when listeners expect a turn-end coming, articulation will start not immediately, but only in the first half cycle after the speaker finished.

To summarize, a long tradition of interaction research has identified a set of prosodic features that typically coincide with turn-ends. It has been proposed that these features can warn the listeners that turn is ending soon, so they can prepare to take the floor.

1.2.3.2. Predictions of turn-length

In their seminal paper, Sacks, Schegloff and Jefferson (1974) suggested that interactants use the syntactic frame to predict or "project" (as they prefer) the overall structure of the incoming turn, and thereby predict when a turn is going to end:

"There are various unit-types with which a speaker may set out to construct a turn. Unit-types for English include sentential, clausal, phrasal, and lexical constructions.

Instances of the unit-types so usable allow a projection of the unit-type under way, and what, roughly, it will take for an instance of that unit-type to be completed" (Sacks et al., 1974: 701).

The authors also acknowledge the role of intonation in the projection of the length of utterances:

"Clearly, in some understanding of 'sound production' (i.e. phonology, intonation etc.), it is also very important to turn-taking organization. For example, discriminations between *what* as a one-word question and as the start of a sentential (or clausal or phrasal) construction are made not syntactically, but intonationally." (Sacks et al., 1974: 721-722).

Prediction of the constructions which build up turns has been studied less extensively than turn-yielding cues. However, the proposal from Sacks et al. suggests that listeners can predict whether a turn is, for example, one word long or a multi-unit construction. Such predictions help the listeners to predict the turn-end. Although Sacks et al.'s turn-taking paper does not mention lexical information, prediction of words contained in a turn might also help listeners to anticipate the length of turns. If listeners can predict whether a one, two or three-syllable word will finish the turn, that is, they might also predict whether the turn will continue for about 200, 400 or 600 milliseconds. Listeners might also take into account speech-rate in estimating the duration of predicted words and constructions.

Although turn-yielding cues have been typically found in prosody, and prediction of turn length could be related to lexical and syntactic information, these two mechanisms do not necessarily differ in the type of information they are based on. Features in prosody, for instance, might also predict whether the turn will be continued by one, a few or by many words (i.e. they might predict how long a turn will continue; see the next section). In addition, lexical information like turn-final tag questions could also express whether a turn is ending soon without providing further information about turn length, and may thus function like turn-yielding cues. For example, Local and colleagues (1985) have shown that *you know,* together with certain prosodic features, can be interpreted as a turn-yielding cue in London Jamaican English.

Hence, the main difference between the accounts emphasizing turn-yielding cues versus prediction of turn-length is the type of information provided. On the turn-yielding account, turn-final cues are assumed to signal imminent turn-ending, meaning listeners can not anticipate the turn-end earlier. In contrast with this, prediction of turn-length could inform the listeners about how long a turn might last, and predictions can be adjusted online according to the input. Interactants can thereby anticipate the turn-end early during the turn, and over the course of the turn's gradual production.

1.2.3.2.1. Experimental evidence for prediction of turn-length

One of the few attempts to study predictions of turn-length focused on the role of prosody. Grosjean (1983) studied whether listeners can predict how long a sentence will last based on the prosodic information in sentence-beginnings in English. Participants listened to the beginnings of sentences and they were asked to guess with how many words the sentences continued. In the segments they heard, the semantic and syntactic information was identical, but the prosody was different. The unheard part of the sentences continued either with zero (i.e. the segment simply ended), one, two or three increments. Results showed that participants could accurately predict the length of the sentences if they heard some part of the last word of the played segment (i.e. the first word that was potentially last, and not before). Grosjean concluded that prosodic information becomes available for the prediction of sentence length only when semantic and syntactic information cannot help further.

Another study used the same experimental task with French sentences (Grosjean & Hirt, 1996). In contrast with the earlier study (Grosjean, 1983), listeners could only differentiate whether sentences would continue or not, but could not accurately predict the length of sentences. Both of these studies relied on sentences read aloud, and so it remains an open question as to whether the prosody of utterances in spontaneous speech contains long-range information about utterance length.

Turn-final prosody has been implicated in turn-yielding cues (see section 1.2.3.1). It is not known, however, whether these prosodic features also provide precise temporal information about the length of the final word. Were such information available in turn-final prosody, listeners could probably estimate turn-ends more precisely, especially if combined with speech rate. Regardless, such estimations provide information about the turn-length only at a point very close to the end of the other's turn.

Regarding the use of lexical and syntactic information, experimental studies have rarely examined whether such information is used to make predictions about turn length. De Ruiter and colleagues (2006) suggested that lexico-syntactic information plays a major role in the timing of turn-taking based on experimental results (see 1.2.3.1), but they did not explain how it helps in turn-end predictions. Although lexico-syntactic information is a prime candidate for providing information about turn length, grammatical material at the ends of turns (e.g. tag questions) may also be used as turn-yielding cues. Thus, there is not yet any experimental evidence clearly showing that listeners can anticipate turn-endings based on lexico-syntactic information.

1.3 The framework for the thesis

1.3.1 Temporal estimation based on syntactic and lexical information

We saw that listeners make predictions at different levels of language comprehension, including predictions of the phonological forms of upcoming words (1.2.1). We also reviewed accounts indicating that listeners predict the moment when conversational turns end (1.2.3). We also saw two possible mechanisms for how listeners could predict turn-ends: (1) by signals in speech which warn the listener the turn is ending soon, or (2) by predicting the length of the upcoming part of the turn. The dissertation focuses on this latter mechanism. More precisely, most of the studies in the dissertation examine whether the prediction of words and syntactic phrases allow the listener to predict the length of conversational turns. It was discussed earlier that prosodic information might help listeners predict how long a turn will be (1.2.3.1). The studies in this dissertation, however, focus not on prosody, but on the role of the lexical and syntactic information. Prediction of turn-length based on syntactic frames and individual words in turns could provide an economical account of the interaction between language production and comprehension in conversation. Syntactic and lexical predictions could help listeners and facilitate the speed of preparation for the next turn in two ways:

(1) Prediction of turn content enables listeners (next speakers) to start to prepare an appropriate response before the end of the current turn; and
(2) Next speakers can estimate when the current turn will end.

With regard to (1), note that it might not be necessary to predict the actual words of turns in order to able to produce a relevant answer. The prediction of the overall content of a turn (i.e. its social action) might be sufficient to initiate preparation of the next turn (1.2.3). Nevertheless, the predicted syntactic frames and words could further facilitate the correct understanding of the turn. As already reviewed (1.2.1), word prediction studies have mostly used controlled linguistic materials with strong contextual constraints. Therefore, an important question of this thesis is whether the initial part of a conversational turn can provide sufficient contextual information for predicting the final words of that turn (see 1.4.1).

With regard to (2), the role of syntactic phrases and lexical information has been little studied with respect to prediction of turn-length. Therefore, the next research question is whether correct predictions of the (syntactic and lexical) content of turns correlate with better predictions of the turn-ends measured by button-press or by verbal responses. Furthermore, oscillatory EEG correlates associated with predictions of the turn endings are also investigated. The EEG study examines how early during a turn expectations arise about the timing of the turn-end (see 1.4.1).

A fundamental property of the human brain is its capability to make predictions (Bar, 2009; Friston, 2010). These predictions provide information not only about what is going to happen, but also about when it is going to happen (Buhusi & Meck, 2005; Nobre, Correa & Coull, 2007). During conversation we engage in a joint activity with others (Clark, 1996), where these predictions are extremely important. In joint activity, participants need to predict what and when the other is going to do in order to produce smooth coordination (Sebantz, Bekkering & Knoblich, 2006). Hence, it is an intriguing question whether prediction of what the other is going to say helps in the prediction of how long the other is going to speak. In other words, the question is whether the predicted linguistic information also provides information about its duration.

We do not know the exact units (or precision) of temporal predictions of linguistic information. The studies in this dissertation focus on temporal information associated with the prediction of words, the number of words and the number of syllables of words. These predicted elements are assumed to provide information about their duration.

1.3.2 Possible mechanisms of how prediction of turn-length could lead to fast turn-transition times

The starting point of the dissertation is the timing of turns during the course of natural conversations (see 1.1). The previous section posed the question of whether the prediction of a turn's linguistic content helps estimate turn duration. A related question was whether these temporal estimations enable speakers to fit their turn to the end of the previous turn. So far, however, there has been little discussion of why and how predictions of turn could facilitate the timing of speech.

Turn-transition times are often short in natural conversations despite the relatively long speech preparation process (1.2.2). One possible explanation is that speakers of a next turn time the start of the speech preparation (from lexical selection until articulation) relative to the estimated end of the current turn. For example, if speech preparation takes 600 ms long, next speakers initiate the speech preparation 600 ms before the estimated turn-end. Another possibility is that the initial stages of the speech preparation occur as soon as the content of the current turn is predictable and a response can be conceptualized. In this view, listeners may have to delay starting responses so as not to overlap with the current speaker, meaning they would only start speaking once they have enough evidence that the current turn is ending soon. The initiation of the articulation would take less than 600 ms, because the content of the to-be-produced turn has been already prepared. Furthermore, the anticipation of the turn-end could also facilitate the speed of articulation via attentional and motor preparation processes. Therefore, a further research question of this dissertation is whether speakers start speech preparation relative to the estimated turn ending or, alternatively, as soon as they can predict the content of the current turn (1.4.2.).

1.4 Outline of the thesis

1.4.1 Prediction of the content of turns and the prediction of the turn-end

The first research question of the dissertation was whether the initial part of a conversational turn can provide sufficient contextual information for predicting the final words of that turn. Hence, chapter 2 and chapter 4 present studies which examine whether final words of conversational turns can be predicted. It was also asked whether correct predictions of the (syntactic and lexical) content of turns correlate with better predictions of turn-endings. The prediction of turn-endings is studied by a button-press paradigm in chapter 2 and chapter 4. In these two studies, participants who are listening to conversational turns are asked to press a button when they think the turn ends. Chapter 3 employs a question-answer paradigm in order to study turn-end predictions by verbal responses. In this experiment, participants are asked to answer questions, and the speed of the verbal responses to the questions (i.e. turn-transition times) are measured. Oscillatory EEG correlates associated with predictions of the turn endings are also examined in chapter 4. This study focuses on the third research question. It examines how early during a turn expectations arise about the timing of the turn-end.

The stimuli and button-press results of an earlier study (De Ruiter et al., 2006) were used for the experiment in chapter 2. In that study, participants were presented with recordings of single turns taken out of conversational context, and were instructed to press a button exactly when the turns ended. Hence, for each turn we knew how precisely listeners could predict when they end. We tested whether turns with better button-press results (i.e. closer to the turn end) were associated with better predictions regarding the last words of turns, and whether they were associated with better predictions regarding how many words will finish the turn. For this, we used a gating paradigm where we presented the initial fragment of turns to participants whose task was to guess how the turns might continue. Our prediction was that better word predictions correlate with better button-press results.

Chapter 3 employed a question-answer paradigm in which participants were asked to answer questions about images on a computer screen. For half of the questions, they could answer the questions before the question ended; for the other half, they could guess the correct answer only upon hearing the last word of the question. The second experimental manipulation controlled for whether participants could guess the length of the last word of the questions in advance. We asked the participants to answer the question as soon as the questions ended and we measured the response times. We predicted that speakers would answer the questions faster if they could predict the answer earlier. Similarly, we predicted that participants would answer faster if they could predict the length of the last word of the questions in advance.

Chapter 4 describes an EEG study resembling the experimental task in chapter 2. First, we tested how predictable the last few words of conversational turns are. Then, we asked another group of participants to listen to these turns and to press a button exactly when the turns ended. We measured the button-press latencies and the EEG signal before the button-press. Regarding the behavioral results, our prediction was that button-presses were more precise when the last words of the turns were predictable. Regarding the EEG results, we were interested how early during the turns we could find differences in the oscillatory dynamics between turns with predictable and less predictable endings.

1.4.2 The timing of the speech preparation process

The final research question of the dissertation concerns the nature of the anticipatory mechanism that helps speakers time their turns closely to the end of the previous turn. In section 1.3.2, two possible mechanisms of timing of turns were mentioned: (1) speakers time the speech preparation process relative to the turn-end, or (2) speakers start the speech preparation process as soon as they can conceive of an answer. Hence, the study in chapter 5 examines whether speech preparation is timed relative to the predicted moment of articulation.

In this study, we presented non-words to participants and asked them to say these words after an auditory stimulus (beep). Experimental trials were presented in the two blocks, with beep duration differing between the two. Within each block, the beep duration was the same, participants could predict the end of the beeps in that block. If participants start speech preparation relative to the predicted beep-ending, their speech preparation would be delayed with longer beeps. We analyzed the oscillatory dynamics of the EEG after the beeps started. We expected to find differences in the EEG if speakers delayed speech preparation in the longer beeps condition.

CHAPTER 2

2 PREDICTION OF TURN-ENDS BASED ON ANTICIPATION OF UPCOMING WORDS

Based on:
Magyari, L., De Ruiter, J.P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, 3(76), doi: 10.3389/fpsyg.2012.00376.

2.1 Abstract

During conversation listeners have to perform several tasks simultaneously. They have to comprehend their interlocutor's turn, while also having to prepare their own next turn. Moreover a careful analysis of the timing of natural conversation reveals that next speakers also time their turns very precisely. This is possible only if listeners can predict accurately when the speaker's turn is going to end. But how are people able to predict when a turn-ends? We propose that people know when a turn-ends, because they know how it ends. We conducted a gating study to examine if better turn-end predictions coincide with more accurate anticipation of the last words of a turn. We used turns from an earlier button-press experiment where people had to press a button exactly when a turn ended. We show that the proportion of correct guesses in our experiment is higher when a turn's end was estimated better in the button-press experiment. When people were too late in their anticipation in the button-press experiment, they also anticipated more words in our gating study. We conclude that people made predictions in advance about the upcoming content of a turn and used this prediction to estimate the duration of the turn. We suggest an economical model of turn-end anticipation that is based on anticipation of words and syntactic frames in comprehension.

2.2 Introduction

We use language most frequently in an informal, conversational setting. Despite this, most of the studies of language comprehension and production are based on experiments that are conducted in a laboratory with highly controlled input, with single subjects. When one leaves the laboratory and takes a closer look at natural conversations, a striking feature emerges that has rarely been investigated experimentally. People are remarkably fast and accurate in switching between listener and speaker roles during conversations. In Dutch conversations, almost half of all turn-taking role transitions take place with a temporal offset of between −250 and +250 ms measured from the end of the current turn (De Ruiter, Mitterer & Enfield, 2006). Such rapid turn-taking is not specific for Dutch conversations, but has been shown to be universal across cultures (Stivers et al., 2009). Yet, most of the recent models of language comprehension and production (for example, Indefrey & Levelt, 2004; Hagoort, 2005; Pickering & Garrod, 2007; but see recently Pickering & Garrod, 2013) do not explain how the production/comprehension system manages to achieve this highly accurate timing.

Almost four decades ago, Sacks, Schegloff & Jefferson (1974) suggested that it is a normative rule in conversations that participants respond as soon as the current speaker has finished. When there are departures from this rule, the gaps or overlaps are interpreted communicatively. For example, a short silence before a response can be a sign of disagreement in the coming response (Davidson, 1984; Pomerantz, 1984). They also argued that listeners must predict the end of the current turn to properly time their own turn. But from the point of view of the underlying cognitive processes, rapid turn-taking is puzzling. People are required to execute two major cognitive tasks during a conversation: they have to both comprehend one utterance and plan another. The short duration of turn-transitions suggest that comprehension and production must occur in parallel toward the turn-ends. Despite this complex process, every day conversations run smooth and effortlessly. The speed and sensitivity for timing of utterances makes the cognitive processes underlying conversations even more complicated. Next speakers also have to predict when a turn is going to end in order to time their own utterances correctly. How do people execute three major tasks in such a short time and how are people able to predict turn-ends with such accuracy?

Some proposals suggest that speakers produce signals that indicate that they are about to finish their turn (Duncan, 1974; Duncan & Fiske, 1977). Another account assumes that a potential next speaker can anticipate the moment when the current turn is going to end (Sacks et al., 1974). The "signaling" approach identifies cues that usually coincide with the end of turns, for example a certain intonation pattern, a drop in pitch or loudness at the end of phonemic clauses (Duncan, 1974). However, these turn-ending cues probably occur too late for the listener, who after all has to prepare a coherent answer as well. Experimental research on production of words and utterances shows that it requires at least 600 ms or more for the production system to arrive from the message level to articulation (Jescheniak, Schriefers & Hantsch, 2003; Indefrey & Levelt, 2004; Schnurr, Costa & Caramazza, 2006). Therefore, it appears very plausible that listeners must know more than a half a second in advance that a turn is going to end. While the signaling approach does not correspond to recent experimental results on language production, the anticipation account has also not provided a model for how turn-end projection is possible. It has been suggested, for example, that turn-ends can be anticipated in advance by a pitch peak that signals that the next syntactic completion point can be a turn-transition point (Schegloff, 1996). And a more recent experimental study (De Ruiter et al., 2006) has shown that the semantic and syntactic content plays a major role in turn-end predictions. But it has been not studied how the semantic-syntactic content helps in estimating the duration of turn-ends.

A few experimental studies of end-of-turn prediction concentrated mainly on the role of intonation versus the semantic and syntactic content of turns. Grosjean and Hirt (1996) used the gating method to investigate if people can predict when French and English sentences end. They presented sentences auditorly in segments of increasing duration while the subjects had to guess with how many words the fragments would continue in a multiple choice response task. The sentences were either short or they were expanded by optional noun-phrases. Participants could only predict whether the segments continued with three or six more words, when they heard the first potentially last word, i.e., at the first point in the sentence where the sentence could end if it was a short sentence without optional noun-phrases. Grosjean and Hirt concluded that prosodic information in English is made available for the prediction of sentence length but only when the semantic and syntactic information are of no help. This result shows that the semantic and the syntactic information play a role in turn-end anticipation. But the experiment used recordings of read sentences. As spontaneous speech differs from read speech (Levin, Schaffer & Snow, 1982; Esser & Polomski, 1988), it is unclear to what degree their results can be generalized to account for processing of spontaneous speech.

De Ruiter et al. (2006) investigated the contribution of the lexical-syntactic content and intonation in turn-end predictions using recordings of natural conversations. They found that people rely on the lexical content and syntactic information in predicting turn-ends. Subjects listened to individual turns taken from Dutch telephone conversations and were asked to press a button exactly at the moment the turn-ended. The duration between the end of a turn and the button-press (called bias) was measured. In different experimental conditions, the turns were presented naturally (as recorded) or a modified version was played. In one of the conditions, the intonation contour was filtered, in another condition the lexico-syntactic content was removed but the intonational information was left intact. When subjects were listening to the original turns, their button-presses coincided with the turn-ends accurately; the distribution of the button-presses was similar to the distribution of the duration of the turn-transitions in the original conversations. There was no change in accuracy when the intonation contour was filtered, but the performance deteriorated significantly when the words could not be understood, even if the intonational information was still present in those stimuli. De Ruiter and his colleagues concluded that the intonation contour is neither necessary nor sufficient for the prediction of turn-ends. These results suggest that the symbolic (lexico-syntactic) information plays an important role in the prediction of turn-endings.

The results of this experiment correspond to the criticism that intonation cues seem to occur too late to be used for turn-end prediction. However, Casillas and Frank (2013) have found that children can predict turn-transitions best when both intonation and lexical information are available. Moreover, even if intonation is not necessary for turn-end predictions, it still can play a role in turn-taking. For example, it has been suggested that the pitch contour can also serve as a turn-keeping signal before a pause, indicating that despite the pause the turn has not finished yet (Caspers, 2003; De Ruiter et al., 2006). Bögels and Torreira (submitted) have also shown in a button-press experiment similar to de Ruiter et al.'s study that truncated turns ending in a syntactic completion point but lacking an intonational phrase boundary led to significantly delayed button-press times. They suggest that speakers adjust their articulation in reference to both syntactic and intonational completions. On the other hand, how exactly the lexical content and the syntax are helping in turn-end anticipation remains an open issue.

We propose that anticipation of the lexical content and the syntactic information helps in the prediction of the time when a turn-ends. In other words, people know *when* a turn-ends by predicting *how* it ends.

For a long time, predictions were not considered to be part of language processing because they were thought to be inefficient and cognitively demanding. But others have recently argued that predictions can help in speeding up comprehension and disambiguate the noisy linguistic input (Kutas, DeLong, Smith & Bar, 2011). Experimental studies using eye-tracking and electrophysiological techniques revealed that predictions can be made at many linguistic levels during language processing. It has been shown that listeners can anticipate upcoming arguments of verbs (Altmann & Kamide, 1999; Kamide et al., 2003), the gender of words (Wicha, Moreno & Kutas, 2004; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005), and also the upcoming word forms (DeLong, Urbach & Kutas, 2005). Kutas et al. (2011) argue that electrophysiological studies show that word features and word forms get neurally preactivated in highly constraining contexts (DeLong et al., 2005).

Timed turn-transitions require prediction of turn-durations during every day conversation. We are interested if predictions of words and syntactic structures are also made during comprehension of everyday conversations and if these predictions help in predicting the duration of the conversational turns. In real life conversations, the context can also provide further information about the speaker's intention ("message" of the turn) even before any prediction about words or syntactic frames are made. These different types of anticipated information could differentially influence the preparation of the production of next turns. When a message is anticipated, this may provide enough information to start the preparation of a response. When syntactic frames or words are anticipated, this could facilitate the accurate timing of the next turn's production. Perhaps words can only be anticipated when both the message and the syntactic form are clear *(Figure 1)*.

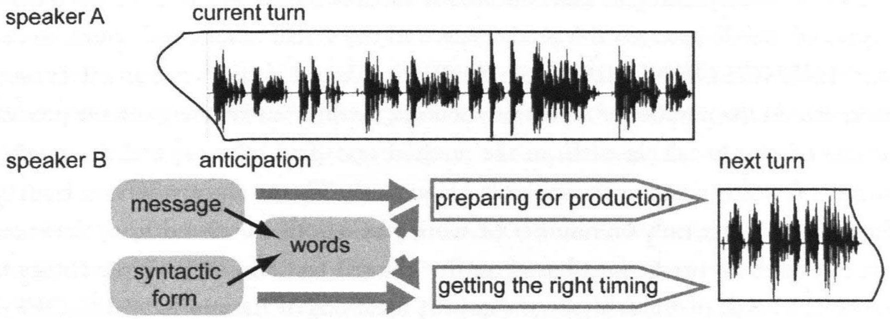


Figure 1. Schematic model of prediction and production processes of the current listener/ next speaker in conversations.

In order to test our hypothesis we conducted a gating experiment using the experimental stimuli of de Ruiter et al.'s (2006) study (these stimuli were single turns from recordings of natural conversations). We presented selected stimuli (single turns) from this study to participants, but cut them off at several points. The participants listened to the turn-fragments or to the entire turn and had to guess how the turn would continue. If they did not make a guess, they had to guess how many words would follow the fragment they heard in a multiple choice task. Our prediction was that accuracy of button-presses to a given turn-end in the earlier experiment correlates with the accuracy with which words the participants guessed follow the presented fragment. We test this prediction by using a mixed-effects model on the guessed words of which prediction is either correct or not (PREDICTION OF WORDS). We test if the accuracy of the button-press responses of the turns (BIAS) has an effect on the proportion of the correct guesses. It is possible that people can predict duration of turns not only by predicting the words coming, but also based on the syntactic structure even when no concrete word predictions are made. We assume that correct syntactic predictions correlate roughly to the expectations about how many number of words follow a fragment. For example, when certain syntactic structure requires obligatory arguments, participants can predict that a certain number of words are still required for the turn being grammatically correct. Therefore, we predict that accuracy of button-presses correlates also with the predictions about how many words follow the fragments. We test this in two ways: (1) First, a mixed-effects model is used on the number of word predictions as binary responses (correct or incorrect number of words is predicted, PREDICTION OF NUMBER OF WORDS). It is tested if the timing accuracy of turns (BIAS) has an effect on the proportion of the correct number of words prediction. (2) Then, another mixed-effects model is used on the number of word prediction as a continuous variable. Here, we code the difference in number of words between the predictions and the actual number of words to come (DIFFERENCE OF NUMBER OF WORDS). We test if BIAS has an effect on this difference. At the number of words predictions, we differentiate between the predicted number of words calculated from the entered text (free guesses) and the predicted number of words given in the multiple choice task. We run the analysis on both type of responses (once only on number of word predictions calculated from the entered text, and once on predictions based on the entered text and on multiple choice task together). In each of the analyses, the cut-off locations of the stumuli (CUT-OFF) are included as a dependent variable besides BIAS.

2.3 Materials and methods

2.3.1 Participants

Fifty native speakers of Dutch (forty-two women and eight men, aged between eighteen and twenty-nine) participated in the experiment. The data from one subject was excluded because the results indicated that he did not understand the task correctly. The subjects were paid for their participation.

2.3.2 Stimulus material

The experimental materials were selected from stimuli used by de Ruiter et al. (2006). In de Ruiter et al.'s study subjects listened to individual turns taken from Dutch telephone conversations and were asked to press a button exactly at the moment the turn-ended. We selected our stimuli material from those turns. It was known for each turn from the results of the earlier study how accurately subjects could predict the end of turns by button-press. We took this information into account in our selection. We used the value of bias at each turn that was calculated in the earlier study (De Ruiter et al., 2006) based on the subject's responses. Bias is the temporal offset between the end of the turn and the button-presses. In de Ruiter et al.'s study, subjects did not react on the occurrence of a stimulus but subjects were trying to press the button exactly at the occurrence of the turn-end. This could result also in "early" responses that occur before turn-end. Therefore, subject's button-press responses are called bias instead of "reaction time." When bias is negative, the subject pressed the button before the end of the turn, when bias is positive the subject pressed the button after the turn. The averaged bias of a turn indicates how accurately subjects could on average predict the time when the turn-ended. A turn with a highly positive bias indicates that subjects pressed the button considerably after the turn-ended. A low bias (small positive value or with a small negative value) shows that subjects pressed the button on time or a little earlier than the turn-ended in average. We used this average bias calculated for each turn to select our stimuli.

For the purposes of the present study, 20 turns with averaged biases ranging from low to high were selected from 10 different speakers (min = −18 ms, max = 330 ms, mean = 159 ms). It was observed in the earlier study that turns with longer duration tend to have a smaller bias. In this study, we were interested if there is a relationship between bias and predictability of the last words of turns, therefore, we tried to avoid differences between turns with different bias caused by differences in the duration of the turns, in the duration of the fragments of turns to be predicted, in the number of words, or in the number of syllables to be predicted. Therefore, we selected turns using the following procedure: Initially, we selected pairs of turns having approximately the same duration, but with one turn with high and one with lower bias. For each turn, four versions were made by cutting off the speech at four different temporal locations *(Figure 2)*.

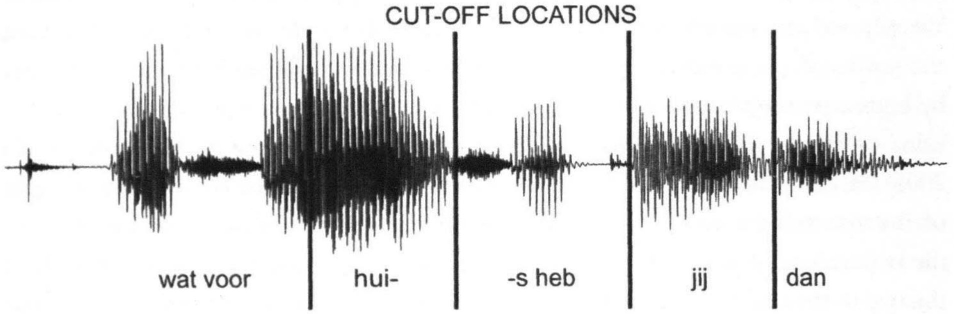


Figure 2. *Sound wave and content of an example turn with four cut-off locations. The audio recording of each turn was cut at four different points (vertical lines) before the turn-end.*

The cut-off locations within each pair were at the same points in time measured from the end of the recordings, but they were different across stimuli pairs. Locations of the cut-off points were determined in each pair according to the boundaries of the two last words of each of the pairs. The turn-pairs with approximately similar duration but with different bias ensured that the bias of the selected turns did not correlate with their duration (r = −0.05, p = 0.82). The pairing of turns were not used further in the analysis of the results because we used bias as a continuous variable but the cut-off procedure based on the pairs ensured that there was no systematic (linear) relation between the bias of the turns and other features of the fragments that had to be predicted. At each cut-off location (from the longest fragment to the shortest across turns), there was no correlation between the bias of the turns and the duration of the cut-off fragments (cut-off 1: r = 0.55, p = 0.14, cut-off 2: r = 0.03, p = 0.89, cut-off 3: r = −0.12, p = 0.63, cut-off 4: r = −0.32, p = 0.22), the number of words to be predicted1 (cut-off 1: r = −0.06, p = 0.81, cut-off 2: r = 0.03, p = 0.88, cut-off 3: r = −0.07, p = 0.76, cut-off 4: r = −0.1, p = 0.71), the number of syllables to be predicted2 (cut-off 1: r = 0.04, p = 0.85, cut-off 2: r = 0.06, p = 0.79, cut-off 3: r = −0.04, p = 0.86, cut-off 4: r = −0.01, p = 0.97), whether the cut-offs were at word-boundaries or not (cut-off 1: r = −0.17, p = 0.48, cut-off 2: r = 0.08, p = 0.75, cut-off 3: r = −0.18, p = 0.44, cut-off 4: r = −0.1, p = 0.71), and average frequency of turns to be predicted (cut-off 1: r = −0.01, p = 0.95, cut-off 2: r = 0.39, p = 0.1, cut-off 3: r = 0.19, p = 0.46, cut-off 4: r = 0.34, p = 0.21). Word frequency was based on the log lemma frequencies of the CELEX database (Webcelex, 2001) of the Max Planck Institute for Psycholinguistics. The frequency of two words which were not found in CELEX was set to 0. As a result of the cut-off procedure each turn had four versions with increasing cut-off fragments from the end, but the exact duration of the cut-off fragments at each location varied across turns. Table 1 shows the minimum, maximum, and average duration of the cut-off fragments (i.e., the duration from the cut-off location to the end of the turn) at all four locations. The duration of the entire (non-cut) turns varied between 1.06 and 2.04 s with mean of 1.48 s.

	Duration (cut-off from end) (in ms)			Number of words to predict			Number of syllables to predict		
cut-off	min	Max	Mean	min	Max	mean	min	max	mean
(shortest) fourth	73	355	195	0.5	1	0.794	0	2	0.882
third	146	641	343	0.5	3	1.2	0	4	1.7
second	217	821	470	1	3.5	1.8	1	5	2.4
first (longest)	361	1237	711	2	5	2.875	2	9	4

Table 1. *Duration, number of words and number of syllables of the cut-off fragments across turns*

Table 2 shows two turns and their English translation with the locations of the cut-offs. Vertical lines indicate where the recordings were cut in the different versions. Notice, that for each turn, two of the cut-off locations were (1) at the word before the last word and (2) before the last word.

Cut-off:	1	2	3	4
Maar dat hoor ik wel via	de	mi-	crof-	oon
but I hear that well through	*the*	*mi-*	*croph-*	*one*
wat voor	hui-	s heb	jij	dan
what kind of	*hou-*	*se do*	*you have*	*then*

Table 2. *Two examples of the experimental stimuli with cut-off locations*

2.3.3 Experimental design

Subjects were randomly assigned to one of five experimental lists. The stimuli in the lists were presented in random order to each subject. Their task was to indicate (on a computer terminal) if the presented segment constituted a complete turn. If the subjects decided that the turn was not complete, they were asked to guess and enter the text they believed would complete the turn. If they were unable to guess how the turn continued then they were presented with a multiple choice task. They were asked to guess with how many words the turn would continue. The subjects were allowed to choose between the following options; (A) one word, (B) two words, or (C) three or more words. Subjects were also asked how certain they were of their responses. They had to indicate this on a four-point Likert scale.

2.3.4 Procedure

Subjects were seated in front of a computer screen and a keyboard with headphones. The instructions were visually presented on the screen. Before each stimulus a sentence was presented on the screen in Dutch, saying: "When you press the space bar you can listen to the next sound fragment two times." Five hundred milliseconds after pressing the space bar, a stimulus was presented two times, with a 1500 ms pause between the two presentations. After the stimulus presentation, the subjects were shown a prompt (>:) on the screen where they were required to type their guess about the continuation of the fragment. If they thought the turn that they were listening to was complete, they had to type: ".". If they were unable to guess how the turn continued, but they did not think that the turn had finished, they were asked to type a "-". When they were unable to guess about the continuation, they were presented with the multiple choice task of number of words predictions. After reading the instructions, the participants did a training session. During the training four stimuli were presented which were not parts of the experimental lists. After the training session, which could include verbal clarifications, the experimenter left the room and the participants could continue the experiment alone.

2.3.5 Data-coding

Two variables were created based on the responses entered. The variable called PREDICTION OF WORDS coded whether the prediction of the words in the unheard part of the turn was correct or not. It was set to 1 when the continuation of the turn was entirely correct. We regarded a response entirely correct when the guess exactly matched the continuations. Usage of synonyms or words from the same syntactic category did not count as a correct response. The variable was also set to 1 when it was indicated correctly that the turn has ended. PREDICTION OF WORDS was set to 0 when the typed-in continuation was incorrect (different words, or more, or less words were guessed) or participants were unable to provide a guess.

The DIFFERENCE OF NUMBER OF WORDS variable represented the difference between the number of guessed words and the number of words actually completing the turn. This value was calculated from the number of words that were entered or if no guess was provided, from the estimation of how many words would complete the turn in the multiple choice task. In this task, the maximum number of words that could be chosen was "three or more words." Therefore, when the difference was more than plus or minus three words, the value of DIFFERENCE OF NUMBER OF WORDS remained plus or minus three. So, the values of this variable ranged between −3 and 3. When the exact difference in the number of words could not be clearly identified (for example, when two words remained to complete the turn, and the participant chose the option "more than three words"), a value with the smallest possible difference was given (in this case, +1).

2.3.6 Statistical analysis

Some responses (less than 1% of all responses) that were not clear (e.g., words that do not exist in the Dutch language were typed in) were excluded from the analysis. It was assumed that recognition that a turn has ended and prediction of words that continue a fragment are different types of tasks. Therefore, the responses given at the full turns were analyzed separately from the responses given at fragments of turns. It was also checked whether the proportion of answers for the different analyses were significantly different from each other using a proportion test.

The results were analyzed using a Generalized Linear Mixed Model (GLMM) for the binary response variable (PREDICTION OF WORDS) and a Linear Mixed Model (LMM) for continuous response variable (DIFFERENCE OF NUMBER OF WORDS) with Restricted Maximum Likelihood (Baayen, 2008; Jaeger, 2008). The analysis was performed with the lmer function of the lme4 package (Bates & Maechler, 2009) in R Development Core Team (2009). For the GLMM, binomial error structure and a logit link function was specified. At each model, we included BIAS, CUT-OFF, and their two-way interaction as fixed variables. The variable BIAS contained the average temporal offset between a turn-end and the button-presses from de Ruiter et al.'s study, CUT-OFF was an index of a cut-off location in a turn ranging from 1 (longest cut-off) until 4 (shortest cut-off, closest to turn-end). Initially, GENDER, AGE, and

ORDER (order in which trials were presented) were included as fixed variables that could confound the results. We included SUBJECT (subject's ID) as random effect. The model simplification worked in the following way: When the three confound variables were not significant, they were removed from the model. Then, when the two-way interaction was also not significant it was also removed. Once a final model was reached, the model was computed again with the lmer function but using Maximum Likelihood and compared to a null model comprising only the random effect. For the model comparison, the R-function ANOVA (Crawley, 2007) that applies a chi-square test was used. When the final model was significantly different from a null model, the p-values of the estimates were examined. For GLMM, p-values of the coefficients were derived from the p-values provided by the output of the lmer function. For LMM, p-values were derived using pvals.fnc R-function that estimates p-values based on Markov chain Monte Carlo sampling (with 1000 samples). Significant interaction effects were further analyzed with using a linear regression model. The model was also evaluated with the ANOVA function that shows by an F-test if the independent variable contributes significantly in explaining the variance in the dependent variable (Baayen, 2008).

2.4 Results

2.4.1 Recognition of turn-ends

In 92% of the cases ($n = 196$), it was correctly recognized that a turn-ended. In the GLMM, PREDICTION OF WORDS (correct or not correct recognition of a turn-end) was a binary response variable. CUT-OFF as a fixed factor was not included because only responses at turn-ends were analyzed. In the final model, BIAS was included as a fixed effect and SUBJECT (i.e. participant) as a random effect. This model (including BIAS and SUBJECT) was significantly different from a model containing only the random effect [SUBJECT; chi-square(1) = 4.66, $p = 0.03$]. The estimate of the coefficient of BIAS shows that turn-ends were recognized better when the turns had a higher bias, but this effect did not reach significance ($\beta = 0.01$, $z = 1.86$, $p = 0.06$). Therefore, it is possible that there is no difference in the recognition of turn ends among turns with different bias.

2.4.2 Prediction of the continuations

Responses for fragments and not for full turns were analyzed. We excluded also three fragments that ended so close to the end of the turn that there was no audible information to be guessed. In the final model, BIAS and CUT-OFF were included as fixed effects [chi-square(1) = 87.46, $p < 0.001$]. A model including the two-way interaction between BIAS and CUT-OFF was not significantly different from the model without this interaction [chi-square(1) = 0.54, $p = 0.46$]. BIAS and CUT-OFF had an influence on the proportion of correct responses. When the CUT-OFF ($\beta = 1$, $z = 8.22$, $p < 0.001$) location was closer to the turn-end, the proportion of the correct answers increased. Figure 3 shows the proportion of correct answers at each cut-off location. When BIAS ($\beta = -0.01$, $z = -4.8$, $p < 0.001$) was higher, the proportion of correct answers decreased. Figure 4 shows the proportion of the correct answers for each turn.

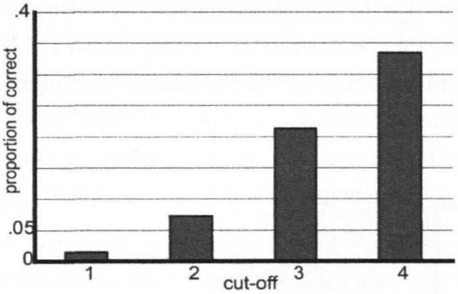


Figure 3. *Proportion of correct answers at each cut-off. As the index of cut-off (x axis, 1-4) increases so it is closer to the turn-end. When the cut-off is closer to turn-end, the proportion of correct answers increases.*

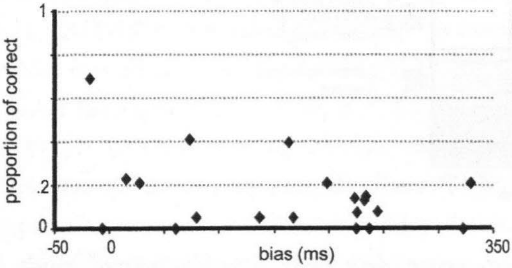


Figure 4. *Proportion of correct answers at each turn. The y-axis shows the proportion of correct answers across cut-off locations. The x-axis shows the bias of each turn. When a turn has a larger bias, the proportion of the correct answers is lower.*

2.4.3 Prediction of the number of words

Our question was also if BIAS predicts the estimation of the correct number of words. We excluded again those items where the full turn was heard. First, we analyzed the number of the predicted words in the free word guesses (in contrast to guesses where they did not type in words but guessed the number of words). The proportion of guesses with correct number of words was significantly larger than the proportion of entirely correct guesses [words = 15%, number of words = 35%, chi-squared(1) = 78.51, $p < 0.001$]. We created a new binary responses variable, PREDICTION OF NUMBER OF WORDS. When the number of words was correctly predicted, PREDICTION OF NUMBER OF WORDS was set to 1, when the number of words was incorrect or no words were entered it was set to 0. We fitted a GLMM using the procedure described in section "Statistical Analysis". None of the confound variables and the interactions showed a significant effect, so the final model contained BIAS and CUT-OFF as fixed effects and SUBJECT as random effect [chi-square(2) = 53.64, p < 0.001]. Figure 5 shows the proportion of correct number predictions calculated from the entered text guess for turn with higher and lower bias. The figure compares the proportion of correct text guesses in terms of number of words collapsing completely correct guesses (solid dark bars) and correct number of words entered as text (solid dark bars and bars with diagonally striped pattern together). The GLMM showed that the proportion of correct number of words increased ($\beta = 0.53$, $z = 7.09$, $p < 0.001$), when the index of CUT-OFF became larger. To sum up, this analysis showed that lower bias did not correlate with predicting the number of words better in the free guesses. But the proportion of correct number of words predictions got higher toward the turn-ends.

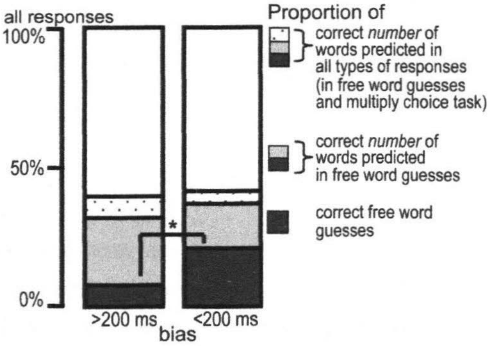


Figure 5. *Proportion of entirely correct guesses (dark bars), correct number of words prediction in the free word guesses (dark bars and light gray bars), and correct number of words predictions in both types of responses, in free word guesses and in the multiple choice task responses (dark, light gray, and dotted bars together). The proportions are shown for turns with lower bias (right bars) and higher bias (left bars). The two groups of turns were created only for illustration. The * shows the significant effect of bias only at the proportion of entirely correct continuations with free word guesses (dark bars). There was no significant effect of bias on the proportion of the correct number of words.*

We also examined if including the number of words estimations in the multiple choice task would change the effect. It was important to examine this because the proportion of all responses where the number of words was correct was significantly different from correct number of word predictions calculated from the entered text responses [words = 35%, all = 41%, chi-squared(1) = 5.26, $p = 0.02$]. In this analysis, PREDICTION OF NUMBER OF WORDS was set to 1 when the number of words was correct based either on the entered text or on the multiple choice number task, and it was 0 when the number of words were not correct. The final model contained BIAS and CUT-OFF [chi-square(2) = 22.81, $p < 0.001$] as main effects. A model with their interaction was not different [chi-square(1) = 0.58, $p = 0.44$]. The effect of BIAS was not significant ($\beta = 0$, $z = 1.49$, $p = 0.14$; Figure 5, dark, stripped, and dotted bars together), but CUT-OFF had an effect ($\beta = 0.32$, $z = 4.59$, $p < 0.001$). When index of CUT-OFF increased, also the proportion of correct number of words (in free guesses and in the multiple choice task together) increased.

We also examined if there was a linear relation between the number of predicted words and BIAS. First, we inspected the number of word predictions among entered text guesses. These were 74% percent of all responses. DIFFERENCE OF NUMBER OF WORDS indicated the difference between the number of words predicted and the number of words in the continuation of turns. When DIFFERENCE OF NUMBER OF WORDS was negative, lower number of words was predicted than the number of words in the continuations. In the final LM model, BIAS and CUT-OFF and their interaction were included [chi-square(3) = 75.18, $p < 0.001$]. The interaction effect was significant ($t = 3.58$, $p < 0.001$), but not the main effects [BIAS: $t = -1.27$, $p = 0.21$, CUT-OFF ($t = 1.16$, $p = 0.25$)]. To examine the interaction, a linear regression model was fitted at each cut-offs. BIAS did not have a significant effect at the first cut-off locations [$F(1) = 0.29$, $p = 0.59$], but it had an effect at all the other cut-offs [2: $F(1) = 10.13$, $p = 0.002$, 3: $F(1) = 10.03$, $p = 0.002$, 4: $F(1) = 31.27$, $p > 0.001$]. At these cut-off locations, when BIAS was higher, the number of predicted words also became higher (2: $\beta = 0.003$, $t = 3.18$, $p = 0.002$, 3: $\beta = 0.003$, $t = 3.17$, $p = 0.002$, 4: $\beta = 0.004$, $t = 5.59$, $p < 0.001$). Figure 6 shows the average difference in number of words predictions at each cut-off locations between turns with higher and lower bias. The relevant data are the comparison of the open squares (turns with lower bias) and the open circles (turns with higher bias) in the figure. It shows that at the first cut-off point, there is hardly any difference between turns with different bias, but a difference emerges at later cut-off points. Please note that the grouping (turns with higher versus lower bias) is done only for visualization in the figure, the statistics was done on BIAS as a continuous variable.

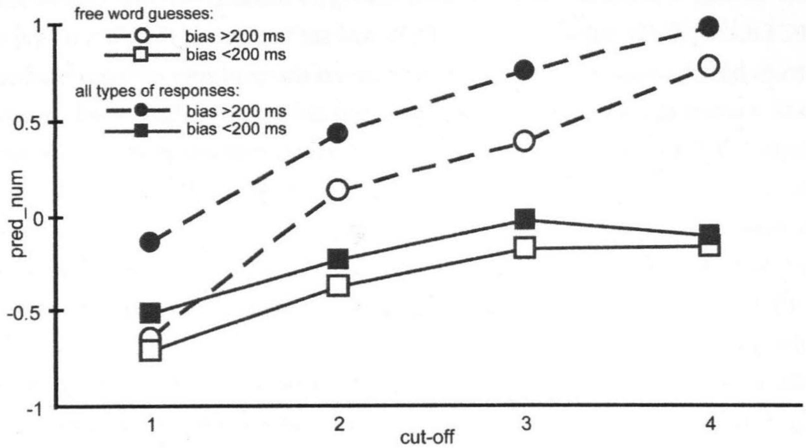


Figure 6. *Average of the difference between the number of words predicted and the number of words which followed a turn segment (DIFFERENCE OF NUMBER OF WORDS) at each cut-off at turn with lower (squares) and with higher (circles) bias among free word guesses (open squares and circles) and among both types of responses, free word guesses and multiple choice task responses (full squares and circles). Both of the analyses (free word guesses and all types of responses) show that the difference in DIFFERENCE OF NUMBER OF WORDS is increasing between the turns with lower and higher bias toward the turn-end. The two groups of turns were created only for illustrating the effect in the figure.*

When predictions of the number of words among all responses (containing either entered text or the multiple choice task) were investigated the results showed the same tendency as the earlier analysis. The final LM model contained DIFFERENCE OF NUMBER OF WORDS as response variable, BIAS and CUT-OFF and their interaction as fixed effects, and SUBJECT as random effect [chi-square(3) = 92.8, $p < 0.001$]. The main effects of BIAS ($\beta = 0$, $t = 0.31$, $p = 0.76$) and CUT-OFF ($\beta = 0.06$, $t = 0.84$, $p = 0.41$) were not significant but their interaction ($\beta = 0.001$, $t = 3.04$, $p = 0.002$) was significant. The linear regression model did not show an effect of BIAS at the first cut-offs [F(1) = 1.68, $p = 0.2$], but it showed an effect at all the other locations [2: F(1) = 12.19, $p < 0.001$, 3: F(1) = 25.69, $p < 0.001$, 4: F(1) = 40, $p < 0.001$]. Figure 6 shows the relevant data in the comparison of the full circles (turns with higher bias) and full squares (turns with lower bias). It shows that the difference between turns is getting larger with lower and higher bias at cut-off locations closer to turn-end. The mean of the number of words predicted (DIFFERENCE OF NUMBER OF

WORDS) was larger than 0 at turns with bias higher than 200 ms [t(319) = 6.05, $p < 0.001$, mean = 0.45], and it was less than 0 at turns with bias lower than 200 ms [t(427) = −4.18, $p < 0.001$, mean = −0.22]. This means that when more words were predicted (compared to how many words is to come) the duration of the turn was estimated to be longer than its real duration. When fewer words were predicted, the duration of the turns were estimated closer to the actual end of the turn by the button-presses.

2.5 Discussion

We examined whether people know when a turn ends because they know how it ends. Therefore, we used a gating method to study how well people can predict the continuation of turn-fragments. Based on an earlier button-press experiment, it was already known how accurately the duration of those turns can be estimated.

The results show that the proportion of correct guesses about the unheard words increased as the cut-off approached turn-ends. The proportion of correct guesses was also higher when a turn-end could be estimated better in time (i.e., it had a lower bias in the earlier button-press experiment). A linear relationship was also found between the bias in the earlier experiment and the difference in number of words people predicted compared to the number of words the fragments continued with. When bias was higher, more words were predicted, when bias was lower, fewer words were predicted.

The results suggest that people make predictions in advance about which words and how many words will follow a partially heard turn, and that they use this prediction in estimating the remaining duration of that turn. Importantly, we show that natural language use is predictable to a certain degree, and we suggest that such predictions are crucial for timed social, verbal interactions. Altogether, the proportion of the correct guesses is not high. However, our criterion for a correct guess was strictly the exact match between the predicted and the coming words. No synonyms or words from the same category were regarded as correct. Moreover, an off-line study perhaps only partially reflects the on-line prediction processes. The turns were also presented without their conversational context. The results also suggest that people follow their prediction in estimation of turn-duration even when those predictions are not entirely correct. This effect has been shown already at the second location of cut-offs that were around 340 ms before turn-ends on average (see Table 1).

One challenge of the anticipatory comprehension account is the explanation of what happens when mispredictions are made. In order to avoid major misunderstandings there must be a monitoring process that compares the actual input to the predicted input (Kolk, Chwilla, van Herten & Oor, 2003; Van de Meerendonk, Indefrey, Chwilla & Kolk, 2011). Interestingly, our results show a correlation between button-presses and the anticipated information. If the actual input is continuously monitored, how is it possible to predict turn-ends based on wrongly anticipated information? And if people follow what they wrongly anticipate, why do they not end up with continuous misunderstanding? We can only speculate on possible answers to these questions, but further work could provide more insight.

When people predicted more words than the number of words that were actually in the turn, button-presses were also late. Late button-presses could have been caused by waiting too long and executing the movement too late, only after noticing the lack of continuation. Late responses also give time for re-planning the production in a conversational situation. In this case, mispredictions may lead to late answers but they do not necessarily lead to misunderstanding or non-relevant responses.

In our data, turns with lower bias (between −18 and 200 ms) were associated with the prediction of a "lower number of words". The button-press results can reflect that when people are predicting fewer words, they prepare for the movement earlier than necessary. A monitoring process can help to stop the movement execution and delay the response until the appropriate moment. Response preparation leads also to faster reaction times (Niemi & Näätänen, 1981). But in other cases, when the language perception system shows that unexpectedly the turn is not ending, it might still be difficult to stop the movement. It has been showed that there is a temporal boundary, a "point-of-no-return" in response preparation and execution (Logan & Cowan, 1984; Sosnik, Shemesh & Abeles, 2007). Ladefoged, Silverstein & Papcun (1973) showed that it is also difficult to interrupt one's own speech especially while one is planning articulation. This could also explain why non-intentional overlaps could occur in conversations. In these cases, early responses may begin with non-relevant or incorrect responses (false starts) but they do not necessarily lead to misunderstanding. It is possible that the speaker has already noticed the misprediction and corrected but could not stop the articulation process. In our stimuli, we did not include turns with very early bias (<−200 ms), however, these also occurred in de Ruiter et al.'s (2006) experiment.

Using intonational cues may seem to be a simpler mechanism for predicting turn-ends but the account presented here is economical in terms of cognitive processing load. It explains how people are able to perform many simultaneous tasks before they start their turn. A next speaker in a conversation has to both comprehend the current turn and formulate and time their own subsequent utterance appropriately. Response preparation takes time and therefore it has to start before the previous turn-ends in order to avoid gaps. Response preparation, however, can be initiated only if the speaker knows roughly how to respond. Therefore, the next speakers have to anticipate not only when a turn ends but also the content of the turn. When the last words of a turn can be anticipated, this provides information about both the content and the duration of the turn. Therefore, anticipatory comprehension is the very same process that helps to formulate the next turn and also to time it properly. This predictive mechanism provides a single, economic solution for the three major cognitive tasks that a listener needs to perform more or less simultaneously: (a) comprehending a turn, (b) preparing to produce the next one, and (c) estimating the end of the current turn. However, further research will be required to reveal the time-course of these processes.

Our data also shows that predictions are made not only about word forms but also about the number of words. We believe that the ability of our participants to predict the number of words reflects their ability to perform syntactic predictions that can help also in estimating turn-durations (see Figure 1 in the Introduction). Timing of turns is also probably influenced by other factors independent of anticipation. Overlaps and delays could also occur with communicative intent and gaps can also occur due to delays in the production or comprehension process of the speaker. But the timing of turns is frequently highly accurate in real life conversations. We suggest that this is possible only when people predict the syntactic form and word forms of turns before the turns end. It is an interesting issue for further research how duration of predicted linguistic elements is represented, for example, how precise duration estimations are and how much these estimations can be influenced by contextual information, for example, by the speaker's speech-rate. For example, Pickering and Garrod's (2013) model of speech comprehension and production suggest that listeners use covert imitation and forward modeling to predict not only the content of the other utterance but also the duration of the speech production process. Gambi and Pickering (2011) argue that these linguistic predictions must be inaccurate because the listeners' forward model is fine tuned to their own production system. However, empirical research still needs to confirm these assumptions.

In our experiment, we showed that people were even able to guess upcoming words and the number of words of turns that were taken out of their conversational context. Button-presses for such "out-of-context" turns were also accurate. In real conversation, the context can facilitate anticipation even further. Listeners not only need to, but are also able to predict the continuation of the speaker's turns before they are completed, and this ability is necessary for engaging in fluent verbal interactions.

CHAPTER 3

3 TEMPORAL PREPARATION FOR SPEAKING IN DIALOGUE

3.1 Abstract

In everyday conversations, the gap between turns of conversational partners is most frequently between 0 and 200 milliseconds. We were interested how speakers achieve such fast transitions. We tested whether speakers already prepare their answers while they listen to questions and whether they can prepare for the time of articulation by anticipating when questions end. In our experiment, participants had to answer questions about images. For some images, it was possible to guess the answer earlier during the questions than for others. In some cases, it was also possible to predict the length of the last word of the questions. The results suggest when listeners know the answer early they start speech production already during the questions. Speakers can also time when to speak by predicting the duration of turns. These temporal predictions can be based on the length of anticipated words and on the overall probability of turn durations.

3.2 Introduction

Recently, there has been increasing interest in interactive language processing in psycholinguistics (e.g. Barr, 2008; Branigan, Catchpole & Pickering, 2010; Pickering & Garrod, 2013; Willems et al., 2010). For a long time, experimental studies of language processing have focused mainly on how individuals comprehend or produce phonemes, words and sentences, while the social setting in which language is used has rarely been investigated. Everyday conversations provide a richer verbal and social context for language comprehension and production than the traditionally used experimental paradigms of psycholinguistic studies. Hence, if we want to study human language capacity an important issue is how an interactional context influences language processing.

One striking aspect of natural conversations is the give and take of turns between conversational partners. Speakers and listeners alternate freely, without the restrictions of any institutional settings (Levinson, 1983). Observations have shown that turn-transitions of natural conversations --the switches between the speaker's and the listener's roles-- happen remarkably quickly. For example, 45% of the turn-transitions are between +250 ms and -250 ms in Dutch telephone-conversations (De Ruiter, Mitterer & Enfield, 2006). Experimental studies have estimated how long the process leading to the initiation of word- and sentence-production takes. This duration is longer than the most frequent turn-transition times of conversations. In picture-naming studies, it takes at least 600 ms to name an object (Levelt, 1989; Indefrey & Levelt, 2004; Indefrey, 2011). Naming times are not radically shorter when words have a higher frequency or when they are repeated during an experiment (Jescheniak & Levelt, 1994; Damian, Vigliocco & Levelt, 2001). In contrast, the mode of the turn-transition durations is shorter than 200 ms in conversations in several languages (Stivers et al., 2009). Given the latency of the speech production process, it seems listeners (who are next to speak) often can not wait until the current speaker finishes, but must begin the production process prior to the end of the current turn (Levinson, 2013).

Therefore, listeners are probably required to execute parallel cognitive tasks for an immediate response. They need to understand the message of the current turn and start to prepare an answer. The short transitions also suggest that next speakers time the production of the answer to the end of the current turn.

In order to prepare an answer before the current turn ends, listeners must know what they want to say. Therefore, it is likely that they anticipate the message of the turn they are listening to. Eye-tracking and EEG studies have shown that people anticipate information at many different levels of language processing. Eye-tracking studies demonstrate semantic predictions made by listeners (e.g. Kamide, Altmann & Haywood, 2003; Altmann & Kamide, 2007; Knoeferle, Croecker, Scheepers & Pickering, 2005). Event-related potential studies show syntactic (Wicha, Moreno & Kutas, 2004; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005) and word-form (DeLong, Urbach & Kutas, 2005; DeLong, Urbach, Groppe & Kutas, 2011) predictions. Kutas, DeLong & Smith (2011) argue for a "strong form" of prediction. EEG studies show that lexical items or their syntactic and semantic features are pre-activated before the predicted item is heard or seen (DeLong et al., 2005; DeLong et al., 2011). They conclude that the anticipation not only facilitates the comprehension of a predicted item when it is encountered in the input, but information can be already processed when it has been predicted. Regarding conversations, this suggests that it is possible to start the preparation of an answer before the turn end, as soon as the content of the turn can be anticipated.

Regarding the timing of a response, the tight transition-times also suggest that next speakers wait for the right moment to deliver the answer (De Ruiter et al., 2006). A timed production is only possible if next speakers know in advance that the current turn will be completed soon. Different mechanisms have been suggested to explain how people know when a turn will end. For example, Sacks, Schegloff and Jefferson (1974) suggested that sentential, clausal, phrasal and lexical constructions can project when a turn will be completed. Although they left open the question how "projection" was accomplished, their account suggests that lexical and syntactic anticipation provides the necessary information for turn-end predictions.

Others have argued that timed production of turns is not a result of anticipation.

This view holds that turn endings are signalled to the listener through a variety of means, including eye-gaze (Kendon, 1967; Duncan, 1974) and changes in prosody (Duncan, 1974; Duncan & Fiske, 1977; Schaffer, 1983; Local, Wells & Sebba, 1985, Local, Kelly & Wells, 1986; Cutler & Pearson, 1986; Beattie, Cutler, & Pearson, 1982; Local & Walker, 2012). Regarding eye-gaze, later studies have shown that turn-transition times are not longer during telephone conversations than in face-to-face interaction (De Ruiter et al., 2006; Stivers et al., 2009; Ten Bosch, Oostdijk & De Ruiter, 2005), that speakers often gaze from the beginning of a question until its end (Rossano, Brown & Levinson, 2009) and that in general, gaze-behaviour is also influenced by the social actions and goals of the participants and not only by the organization of the turns (Rossano, 2012). Therefore, eye-gaze cannot play a systematic role as a turn-transition cue. With respect to prosody, most of the studies have focused on the role of pitch in signalling turn-ends (for an overview see De Ruiter et al., 2006, Local & Walker, 2012). However, Ford and Thompson (1996) showed in their analysis of 198 speaker-changes of face-to-face conversations that speaker changes most frequently occur at points where the syntactic structure, the intonation of the utterance and the conversational action is complete. Therefore, experimental studies are important to evaluate the relative contribution of the different information sources in the prediction of turn-ends. However, only a few experimental studies have addressed this issue (Shaffer, 1983; Beattie, Cutler & Pearson, 1982; Cutler & Pearson, 1986; De Ruiter et al., 2006, Casillas & Frank, 2013). Cutler and Pearson (1986) recorded dialogues by having speakers read written scripts. Fragments of these dialogue-recordings were given to participants who had to indicate whether the fragments were turn-medial or turn-final. They have found that falling intonation contour indicated turn finality while rising intonation served as a turn-yielding cue, but they noted that many of the utterances which were found ambiguous by the listeners also had falling or rising pitch. It appears from a later experimental study (De Ruiter et al., 2006) that accurate turn-end prediction is possible without pitch information and intonation-contour. In this experiment, participants listened to turns taken out of conversational context. The stimuli used in the experiment were from recordings of natural conversations. The task of the participants was to press a button exactly when each turn ended. The button-presses were not less accurate for turns from which the intonation contour was removed compared to the original recordings. But the button-presses were significantly worse when the recordings were manipulated in such way that the intonation was intact but words could not be understood. De Ruiter et al. (2006) have concluded that lexical and syntactic information plays a major role in turn-end predictions.

However, a model of the link between syntactic-semantic information and turn-end predictions is still missing. One possibility is that listeners predict the duration of turns, and therefore, they prepare for the production of a response relative to how late they predict the turn-end coming. Anticipated syntactic structure and words are good candidates for providing information about the duration of the rest of the turn. Magyari and De Ruiter (2012) studied with a gating paradigm whether anticipation of the lexical and syntactic information could help turn-end predictions. Their experiment used a subset of turns from De Ruiter et al.'s (2006) previous experiment. Recordings of the turns were truncated at several points before the end. Participants listened to the segments of the turns and guessed how the turns end. Guesses about the last words were more accurate when initial segments of those turns which had more accurate button-presses in the previous experiment were presented. The number of the guessed words also correlated with the button-presses. Segments of turns with later button-presses were associated with a larger number of guessed words than the actual number of words in the continuations. Magyari and De Ruiter (2012) suggest that anticipation of the syntactic phrase and lexical elements of a turn provides temporal information concerning the turn end.

Another possibility is that the syntactic-semantic content of turns informs the listener whether the currently heard linguistic element will terminate the turn or if it will be followed by further information. In this case, the terminal element serves a cue for response initiation. Speakers do not have expectations of when the turn end will occur, but they will start speech production when they recognize the terminal element (e.g. the last word) of the turn.

The "reaction on a cue" and the "temporal preparation" accounts are not necessarily incompatible with each other. When next speakers are prepared for the motor response, they will respond faster when they encounter a turn-ending cue or the turn-end. They can also inhibit the execution of the articulatory movements by attending to cues which signal that the turn is not ending yet. Such "talk-projecting" cues have already been described in intonational and in non-pitch phonetic features of words (Caspers, 2003; Local & Walker, 2012).

To summarize, we are interested how syntactic-semantic anticipation helps in turn-end predictions. First, confirmation of the effect of linguistic anticipation by an experiment with verbal responses would provide stronger evidence compared to the earlier gating study (Magyari & De Ruiter, 2012). Second, it remains unclear whether syntactic-semantic anticipation facilitates temporal preparation for turn-ends based on the expected duration of the anticipated linguistic information. In other words, we do not know whether listeners anticipate only the semantic and syntactic content of turns or whether they can also anticipate the duration of the predicted semantic and syntactic elements which leads to temporal preparation for the moment to speak.

Studies of temporal preparations mostly employ non-verbal tasks where the time interval between a warning signal and an imperative response signal is manipulated. Several experiments have shown that reaction times to a response signal will decrease if participants can estimate the point in time when the response signal is delivered (relative to the warning signal). When the occurrence of the response signal can be estimated, participants can prepare for their response. Moreover, the level of response preparation is proportional to the subjective probability of the occurrence of the response signal at any given moment. Hence, reaction times will increase when participants are uncertain when the response signal will occur. (Niemi & Näätänen, 1981; Müller-Gethmann, Ulrich & Rinkenauer, 2003; Trillenberg, Verleger, Wascher, Wauschkuhn & Wessel, 2000).

In our study, we also manipulate participant's uncertainty of turn-durations and examine whether this will influence their response times. We are also interested to learn whether speakers start the speech production process before the end of the turn they are listening to.

Thus, our experiment was designed to manipulate preparation of what to respond independently of preparation of when to respond. The participants were tasked with answering recorded questions about objects presented on a display. Regarding the first experimental manipulation (preparation of the content of a response), participants could guess the correct answer either (1) early, i.e. towards the beginning of the question or (2) late, i.e. only when they recognized the last word of the question. With respect to the other manipulation, the two possible last words of the questions were either (1) similar in the number of syllables (either both 3-4 syllables long or both monosyllabic) or (2) different (one possible word-candidate had 3-4 syllables, the other was monosyllabic).

We reasoned that decreased response times when the answer is known in advance, suggest that response preparation starts prior to recognition of the last word of the questions. However, if next speakers start to prepare their answer when they recognize the last word, there should be no difference in the response times. If response time increases when it is more uncertain how many syllables the final word of a question has, it shows word-durations are anticipated and taken into account in response-timing. If next speakers do not predict the duration of words, there will be no difference in response times following this manipulation.

3.3 Methods

3.3.1 Subjects

Forty university students between the ages of 18 and 25 years (10 male and 30 female) participated in the experiment. They were registered in the subject pool of the Max Planck Institute for Psycholinguistics in Nijmegen. All were native Dutch speakers.

3.3.2 Stimulus

Participants looked at pictures on a computer screen and heard questions about the presented pictures. On each picture, black and white drawings of two animals, a tiger and a rabbit and two circles with drawings of colourful objects were presented (see an example in Figure 1). The drawings were from the picture database of the Max Planck Institute for Psycholinguistics or from free web sources.

The questions and names of objects were recorded in advance by a female native speaker of standard Dutch who was asked to read the questions and names of the objects in a natural speaking rate.

3.3.3 Experimental design

An experimental trial consisted of presenting a picture with two scenes (one with a rabbit and one with a tiger above each other) and three questions. Participants were asked to answer the questions while looking at the picture (Figure1). The first two questions were the same in each trial. These questions aimed to control if participants named objects as intended and required participants to observe the picture carefully. These control questions asked which objects each animal has. The tiger and the rabbit were said to "have" the objects in the circle behind them. The third question was the critical question which was different across experimental trials. The critical question always named two objects, and it asked to which animal these objects belonged. The participants' response time was measured only after this last (critical) question.

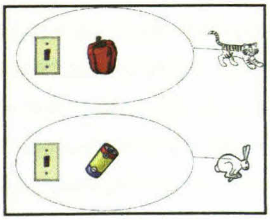
Picture Stimulus	**Audio Stimuli**	**Correct Answers**
	*Welke dingen heeft de tijger?* What kind of objects does the tijger have?	*Een schakelaar en een paprika.* A light-switch and a bellpepper.
	Welke dingen heeft het konijn? What kind of objects does the rabbit have?	*Een schakelaar en een batterij.* A light-switch and a battery.
	Welk dier heeft een schakelaar een bovendien een batterij? Which animal has a light-switch and also a battery?	*Het konijn.* The rabbit.

Figure 1. *An example of the stimuli and the intended answer. An image was displayed on the computer screen (Picture Stimulus, left) while recordings of the three questions were played (Audio Stimuli, middle column). The participants' task was to answer the questions (Correct*

Answers, right column). The experiment was in Dutch. The English translation is written in regular letter type.

The same questions were presented to all participants. However, the pictures belonging to the questions were different in the different experimental conditions. The experimental conditions varied in respect to whether participants could guess the correct answer at the beginning of the critical question or only after recognizing the last word of the critical question (ANSWER: EARLY vs. LATE). The left column of Figure 2 shows two examples of the picture stimuli in the EARLY conditions. When no objects were presented for one of the animals, the participant could be certain that the correct answer for the critical question would be the name of the other animal

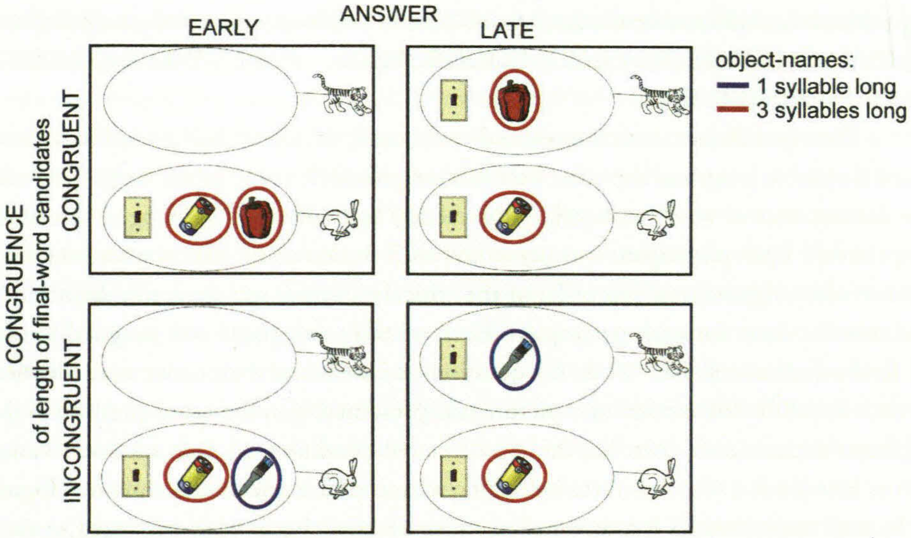


Figure 2. *One set of pictures from the different experimental conditions. The critical question during these pictures and the right answer was the same: "Which animal has a light-switch and also a battery?". When pictures of the left column were presented (ANSWER: EARLY), participants could already guess the answer at the beginning of the question. When pictures were presented in the upper row (CONGRUENCE: CONGRUENT), participants could guess the length of the final word after listening to the first part of the question, because the names of the remaining objects had the same number of syllables (red circles). In the INCONGRUENT condition (lower row) the potential last words had different number of syllables. The red and blue circles were not part of the presented images; they serve only the purpose of illustration here.*

The other experimental manipulation measured if participants could estimate with certainty how long the critical question would take to complete (CONGRUENCE: INCONGRUENT vs CONGRUENT). There were always three different objects presented in each picture stimulus. Two of these objects were mentioned in the critical question. When the first object was mentioned (e.g. "Welk dier heeft een schakelaar …." Which animal has a light-switch…), the final word of the question named one of the two remaining objects (e.g. "paprika" or "batterij", bell-pepper or battery, Figure 1). When the name of these two objects had the same length measured in the number of syllables, participants could more certainly predict the duration of the question before hearing the last word (upper row of Figure 2) (CONGRUENT conditions). When the number of syllables was different between the possible final words, the prediction was less certain (bottom row of Figure 2) (INCONGRUENT conditions). The four experimental conditions were created from these two main experimental manipulations: EARLY-CONGRUENT, LATE-CONGRUENT, EARLY-INCONGRUENT, LATE-INCONGRUENT.

Twenty different critical questions were used, of which half ended in a short word (1 syllable long) and the other half in a long word (3 to 4 syllables long). Similarly, the correct answer was "het konijn" (the rabbit) and "de tijger" (the tiger) ten times respectively. Each participant was presented with these twenty critical questions after the two control questions. The order of the critical questions was randomly determined and was the same for each participant. Each of the participants was assigned to one of four experimental lists. While the questions, answers and their order were identical in each list, different version of a picture was presented with the same question in the different experimental lists. So, the experimental condition of each question varied across lists. Such a set of pictures belonging to one critical question is shown in Figure 2. In each experimental list, an equal number of critical questions belonged to each experimental condition. Each list was assigned to 10 participants.

To summarize, participants were presented with the same 20 critical questions in the same order and 2 control questions before each of the critical questions. Each group of participants was assigned differing conditions for each critical question and each condition was assigned an equal number of times. Additional features were balanced across the critical questions: the length of the last words of the critical questions, the type of the answer ("het konijn" or "de tijger") and the position of the object representing the final word in the visual scenes. A list of the critical questions and the competitors of the final words of these questions can be found in the Appendix (Appendix, Table 2).

3.3.4 Procedure

Participants were seated in a soundproof cabin in front of the screen of a computer presenting the experimental stimuli, wearing headphones, and with a microphone and button-box in front of them. Participants' voices were recorded continuously throughout the experiment using Digital Audio Tape (DAT). Pictures and sound stimuli were presented using the software Presentation 12.1. Button-presses were recorded in log files and on DAT using a beep. The pre-recorded sound stimuli (words and questions) were also recorded when they were played. The sound stimuli and the beep of button presses were recorded on a different track than the participants' voices.

First, a warm-up session was conducted to familiarize participants with the object-images and their names. This session consisted of each object being displayed along with an audio stimulus of its name. This round was run twice. In a third round, the participants were asked to produce the name of the displayed object. Both written and verbal instructions were provided. The round started when the participant pressed a button. Then, each image was displayed for three seconds in a random order along with the audio of its name. In the third run, a trial started when participants pressed a button. First, a cross appeared for 500 ms. Then, the image appeared on screen, the participant produced the name of the object and pressed a button to initiate the next trial.

After the warm-up session, written and verbal instructions were provided for the second part of the experiment. This part started with five practice trials. These trials were similar to the experimental trials but used different stimulus material. Participants were asked to look at the pictures displayed in front of them and answer the three questions about each picture as quickly as possible after the end of the question. With this instruction, we aimed to avoid a quiz-like situation in which participants would answer as soon as they knew the answer. A trial began when the participant pressed a button, and the picture stimulus appeared on the screen. After 250 milliseconds, the first question was played. Once the participants answered, they had to press a button to hear the next question. When all three questions for the picture were answered, a new picture appeared on the screen after a button-press.

3.3.5 Measuring response

The recordings of the audio stimuli and the participants' voice were converted to WAV files using Audacity 1.2.6. The files were analysed with Praat 5.1. Response times were measured as the duration between the end of the critical questions and the beginning of the participant's answer. Negative response times indicate an overlap between question and answer. The ending of a critical question and the beginning of an answer were manually coded using the intensity wave, the spectrogram, and by ear. The beginning of the first speech-sound (excluding inhalations or hesitations) was marked as the beginning of an answer.

3.3.6 Statistical analysis

For statistical evaluation of the results, we used a (linear) mixed-effects model with maximum likelihood estimation. The lme4 package (Bates & Maechler, 2009) of R, an open-source language and environment for statistical computing was used for the statistical analysis (R Development Core Team, 2009). Model-reduction and the selection of the random-effects are described in the Appendix (Statistical Analysis).

We analysed the effect of the control and the experimental variables on the response times. Control variables were variables which could effect response times but were not consistently manipulated in the experiment. These variables were (1) age of the participants (AGE), (2) gender of the participants (GENDER), (3) the order of stimulus presentation (ORDER), (4) the experimental list (LIST), (5) whether the answer contained a determiner ("het" or "de") (DET), (6) whether the answer was "het konijn"/"konijn" or "de tijger"/"tijger" (ANSWER_TYPE). Experimental variables were variables which were manipulated consistently in the experimental design. ANSWER was a factor with two levels which showed whether participants could know the answer for the critical questions early or late, CONGRUENCE was a factor with two levels which coded whether the candidate final words of the critical question were the same or different in length and FINAL_WORD was a continuous variable. FINAL_WORD coded the duration of the last words in milliseconds. It was included because the questions differed in the length of the last words as a result of the experimental design.

We calculated a mixed-effects model with the experimental and control variables as fixed effects and subjects and items as random effects. All continuous variables were centred.

3.4 Results

One participant's data was excluded from the statistical analysis because of remarkably different responses compared to all other subjects (average response time of this subject were more than 2 standard deviations shorter than the mean of all responses). Answers that were wrong, contained hesitations, false starts, laughter were excluded. Answers were also excluded if participants already pressed the button for the next trial while they were answering the critical question. The analysed data had 620 data-points (mean = 336 ms, median = 299 ms, sd=258, min = -427 ms, max = 1442 ms) *(Figure 3)*.

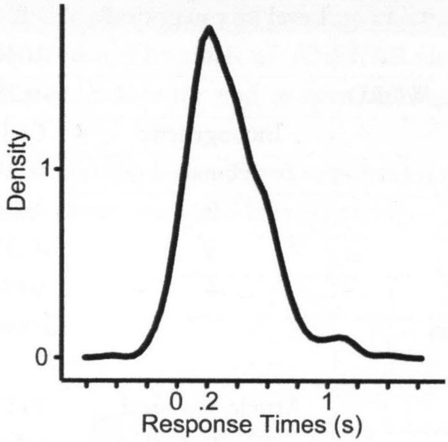


Figure 3. *Density plot of the analysed response times (N=620). X-axis shows the response times in seconds.*

In our mixed-effects model, the response variable was the response time, measured as the duration from the turn end to the beginning of the answer. Fixed effects were whether participants could know the answer early or late (ANSWER), whether the competitor of the final word was the similar or different in length (CONGRUENCE), the duration of the final word of the critical question as a continuous variable (FINAL_WORD) and the control variables (AGE, GENDER, LIST, DET, ANSWER_TYPE). The experimental variables participated in the initial model with all of their interactions; the control variables were included as single main effects. Subject and items were random effects. After model simplification and the expansion of the random effects, the final model contained the main effect of all variables and the interaction of CONGRUENCE and FINAL_WORD. Subjects, items and the slope of ANSWER under subjects were the random effects in the final model. Table 1 summarizes the results.

Variable	Level (if categorical)	β	$X^2(1)$	p
ANSWER	late	0.046	5.51	0.019
Duration of FINAL_WORD		-0.38	28.04	<0.0001
CONGRUENCE	Incongruent	0.01	4.29	0.117
GENDER	Female	0.07	1.32	0.25
Experimental LIST	2	-0.04	0.57	0.904
Experimental LIST	3	-0.02		
Experimental LIST	4	0.01		
ORDER of stimulus presentation		0.0005	0.08	0.771
DET	Article was used	-0.004	0.015	0.903
ANSWER_TYPE	"De Tijger"	0.006	0.084	0.772
FINAL_WORD* CONGRUENCE	Incongruent	-0.24	4.168	0.041

Table 1. *Beta-coefficients, chi-square and p-values of the fixed effects in the final mixed-effects model. Chi-square and p values were obtained by model-comparison*

ANSWER, FINAL_WORD and the interaction of FINAL_WORD and CONGRUENCE showed significant effect on the response times. When the answer was known earlier, participant answered earlier (Figure 4). In the EARLY ANSWER condition the mean of response times was 320 ms and the median was 269 ms, in the LATE ANSWER conditions the mean of the response time was 361 ms and the median was 314 ms.

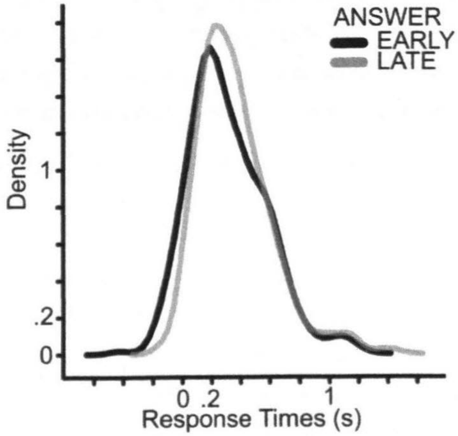


Figure 4. *Density plot of response times in the ANSWER conditions. The dark line shows the densityplot of responses when participants knew the answer for the questions early; the light grey line shows the responses when the participants could guess the answer only at the last word of the question.*

The fastest response in the EARLY condition was around – 400 ms while the shortest question was 3 s long. In the EARLY condition participants knew the answer for the critical question as soon as the picture-stimulus was presented. Therefore, it is unlikely that the effect of ANSWER could arise because participants answered as soon as they knew the answer to the questions. The effect of ANSWER shows that participants started to prepare their answer before the end or even before the last word of the question.

The duration of the final words also effected response times significantly. When the words were shorter, participants answered later (Figure 5).

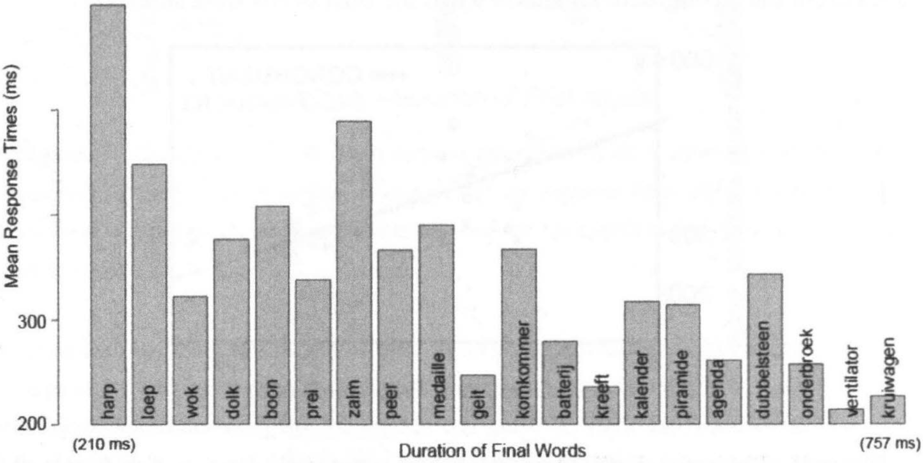


Figure 5. Mean response times for each critical question. The y-axis shows the average of the response times in ms. The mean response times are plotted in increasing order of the final word-duration of the critical questions on the x-axis. The final word of each question is written on the bars. Response times are faster when the final words are longer.

In an additional analysis (see Appendix, Additional analysis), we checked whether the effect of the final words duration arose because participants started the production of their answers as a reaction on phonetic or lexical cues. If such a cue is closer to the turn-end in questions with shorter final words, it could also explain why the response times vary as a function of final word duration. The examined cues were: the beginning of the final word, the beginning of the question, the beginning of the vowel in the final stressed syllable, the peak of the final rise in pitch, the uniqueness point of the final word. According to our analysis, the position of cues relative to turn-ends cannot alone explain the decrease in response times when the final words are shorter. Hence, it is likely that the expected duration of the final words influenced response times.

The interaction between the duration of the final words and CONGRUENCE shows that the duration of the competitor word affected the response differently for shorter and longer final words. Figure 6 shows the mean response times as a function of final word durations in the congruent and incongruent conditions; and the regression lines, respectively. In the congruent condition the number of syllables of the final word and the competitor was the same, in the incongruent condition the number of syllables were different. Our initial hypothesis was that uncertainty about the duration of the critical question (incongruent condition) leads to less preparation which results in longer response times. This has been not confirmed, because CONGRUENCE did not have a significant main effect. However, Figure 6. shows that response time increased in the incongruent condition when the final words were shorter.

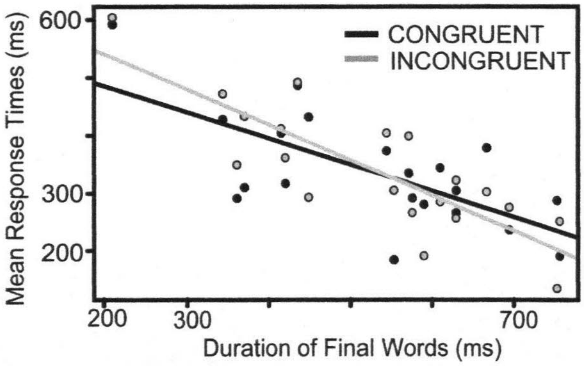


Figure 6. *The average of the response times as a function of the duration of the final words in the congruent (black) and in the incongruent conditions (gray). The final word-durations are presented on the x-axis, mean response times on the y-axis. The gray and black line shows the regression line between mean response times and final word duration in the congruent (black) and in the incongruent (gray) conditions.*

Figure 7. shows the difference of the mean response time in the congruent and in the incongruent conditions for each critical question in the order of the durations of their final words.

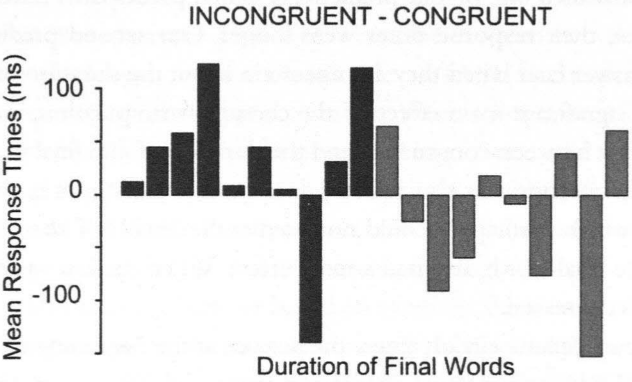


Figure 7. *The difference of the mean response times between the incongruent and congruent condition for each critical question. Y-axis shows the response times differences in ms, the x-axis shows the critical questions in increasing order of the duration of their final words from left to right.*

In a post-hoc analysis, we computed Wilcoxon signed-rank test on the mean difference of response times between the congruent and incongruent conditions separately for the critical questions with the 10 shortest (Figure 7., dark grey) and with the other 10 final words (light gray). We tested whether the difference between conditions was different from 0. For the shortest final words, there was a trend in the deviation from 0 mean (V=45, p=0.084), while there was no trend or significant difference for the other group (V=19, p=0.43). This suggests that uncertainty of the final word duration increased response times only when the final words had a short duration.

3.5 Conclusion

This study measured participants' response times when answering questions. Response time was measured between the question end and the beginning of the answer. We varied in different experimental conditions whether people knew the correct answer early or late and whether they could predict with certainty the number of syllables of the words which ended the questions. The level of certainty was manipulated by presenting questions that could end in the names of one of the two objects in the image. The objects were chosen so their names either have the same (congruent condition) or different number of syllables (incongruent condition). The questions also differed in duration of their final words.

The results confirmed one of our predictions. When participants could only prepare the answer late, their response times were longer. Our second prediction was that participants answer later when they are uncertain about the duration of the question. There was no significant main effect of the certainty-manipulation, but there was an interaction effect between congruence and the duration of the final words. When the final word of a question was shorter, response times were longer in the incongruent conditions in which participant could not predict the length of the final words. The duration of the final words also had a main effect. When the last words were longer, response times decreased.

When participants already knew the answer at the beginning of the questions, they responded in average 310 ms after the question end. This shows that participants did not answer as soon as they knew the answer but were waiting for the right moment to start to speak. The difference in duration of response times in the early and late answer conditions also shows that participants started speech production before the recognition of the last word of the questions in the early condition. The results confirm that next speakers in a conversation will prepare their turn earlier if they know what to respond.

When the last words were longer, participants answered faster. One explanation of this effect could be that participants answered in a fixed time interval relative to a feature in the last words, for example, the beginning of the last word, the uniqueness point, the final pitch or the final stressed syllable. If any of these features occurs earlier relative to question end in longer final words, and participants time the start of the production of their answer to this point, it could explain the shorter gap after longer words. However, in our additional analysis we found that none of these cues alone could explain the difference in response times according to final word duration. Therefore, it is likely that the expected final word durations influenced response time based on the probability distribution of final word durations throughout the experiment. In interval-timing experiments when the momentary probability of the end of an interval is higher, reaction times decrease (Niemi & Nätäänen, 1981). In our experiment, as the duration of the final word continued to grow, the probability of its ending got higher (see Appendix, Figure 11.) Hence, the level of response preparation also increased with the duration of the final words, which is reflected in the response times.

The interaction between congruence and final word durations suggests that temporal expectations independent of duration probabilities also effected the response times. For the critical questions with shorter final words, response times were longer when the length of the competitor was incongruent. This effect was not present for questions with longer final words. This result corresponds to results of studies where a symbolic cue predicts the duration of the interval between a warning and a response signal. When there is a mismatch between the symbolic cue and the actual interval duration, reaction times increase. However, when the duration between the warning and response signal is longer, reaction times are less affected by the mismatch. This is explained by re-orientation of temporal attention. When a response signal is expected to appear early but it appears late, participants still have time to focus their attention on the later time-point (Coull & Nobre, 1998; Capizzi et al., 2012). In our experiment, when the final word did not end early, participants still had enough time to prepare for a later ending independently of our uncertainty manipulation. In non-linguistic tasks, temporal expectations which are based on symbolic features of a warning signal are found to be dependent on the attention towards the feature (Rohenkohl, Coull & Nobre, 2011) and on the working-memory load (Capizzi et al., 2012). Therefore, it is a question for further research whether temporal expectations based on the length of predicted words are attenuated when listeners are facing demanding parallel tasks, for example, difficulties in comprehension or production.

Our results suggests that participants prepared the production of their answers before the question ended and tried to time their production to the end of the questions, yet there was still a substantial gap between the end of the questions and the beginning of the answers. The average duration of response time was 336 ms with a mode between 200 and 400 ms (median=299 ms) (see Figure 3) which is longer than the average duration and mode of turn-transition times reported in Dutch face-to-face conversations (mean=108 ms, mode=100) (Stivers & al., 2009). But this difference is not surprising because the question-answer sequences of our experiments are far from a natural, conversational setting. Moreover, answers which contained hesitations, false starts or laughs were excluded from our analysis, and the beginning of the answers were rigorously coded at the beginning of the first speech-sound of the answer. In everyday conversations, a turn may start with hesitations before the actual answer. Figure 8 shows the frequency of the response times when responses are grouped according to length of the final words (short: monosyllabic or long: polysyllabic) and when the answer was known.

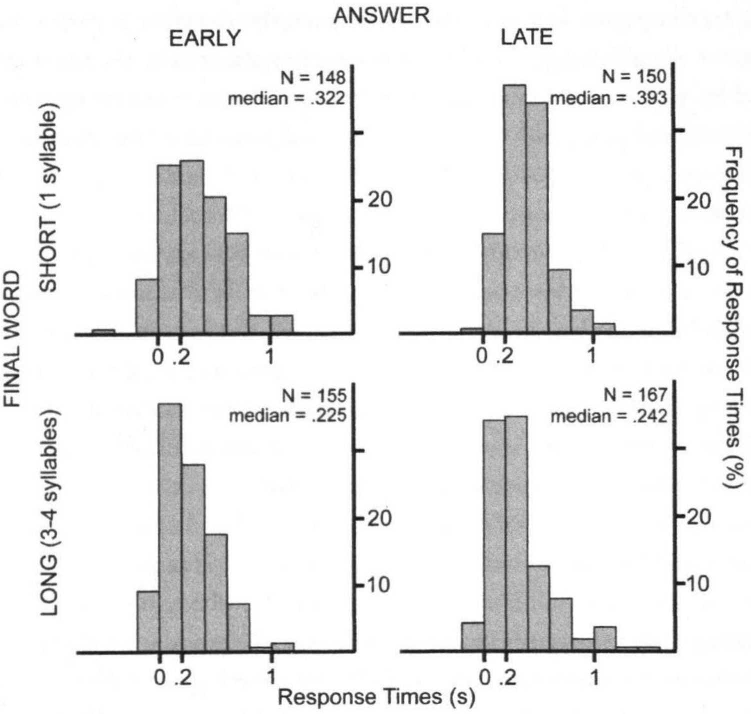


Figure 8. *Histograms show the frequency of response times in 200 ms bins in the different experimental conditions. The upper panels show response times when the final words were monosyllabic (SHORT); the lower panels show response times when the final words were polysyllabic (LONG). The response times presented in the left panels were produced when the answer was known at the beginning of the questions (EARLY), the right panels show results when the answer could be known only at the last word (LATE).*

The medians of response times were around 200 ms at turns with longer final words. When the turns ended with a monosyllabic word the medians were larger than 300 ms even if participants knew the answer already from the beginning of the question. Hence, the experiment introduces a delay in the response times by the experimental manipulation. Therefore, we conclude that the most frequent turn-transition times (0-200 ms) of everyday conversations can be only produced when speakers are prepared for the moment when the preceding turn ends roughly before the last syllable.

To summarize, we showed that speakers will start to prepare for the production of their answer before the current turn ends and they also prepare for the time to speak. This preparation can be influenced by linguistic and non-linguistic temporal expectations.

3.6 Appendix

3.6.1 Stimuli Material

Critical Questions	Competitors of the final words	
	Short	Long
Welk dier heeft een casette en bovendien ook een loep?	spons	antenne
Welk dier heeft een magnetron en bovendien ook een wok?	stronk	kandelaar
Welk dier heeft een lucifer en bovendien ook een kruiwagen?	pauw	kruiwagen
Welk dier heeft een kastanje en bovendien ook een Piramide?	schaar	verrekijker
Welk dier heeft een torpedo en bovendien ook een dobbelsteen?	vlecht	aansteker
Welk dier heeft een brievenbus en bovendien ook een geit?	kruk	paddestoel
Welk dier heeft een liniaal en bovendien ook een harp?	zeis	ananas
Welk dier heeft een boterham en bovendien ook een onderbroek?	taart	boterham
Welk dier heeft een accordeon en bovendien ook een Kalender?	spin	paraplu
Welk dier heeft een zonnebril en bovendien ook een Medaille?	kam	diamant
Welk dier heeft een sigaret en bovendien ook een boon?	kaars	piano
Welk dier heeft een schakelaar en bovendien ook een batterij?	kwast	paprika
Welk dier heeft een portemonaie en bovendien ook een komkommer?	fluit	microscoop
Welk dier heeft een computer en bovendien ook een peer?	sjaal	helikopter
Welk dier heeft een aubergine en bovendien ook een prei?	tol	bikini
Welk dier heeft een asperge en bovendien ook een zalm?	kurk	pantoffel
Welk dier heeft een telefoon en bovendien ook een agenda?	bijl	envelope
Welk dier heeft een capuchon en bovendien ook een ventilator?	zaag	thermometer
Welk dier heeft een camera en bovendien ook een dolk?	kers	parasol
Welk dier heeft een stofzuiger en bovendien ook een kreeft?	tang	parachute

Table 2. *The critical questions and the competitors of their final words listed in the order of their presentation.*

3.6.2 Statistical Analysis

3.6.2.1 Model-reduction and random-effect structure in the mixed-effects models

Model-reduction and the selection of the random-effect structure of the mixed-effects analysis were done in three steps. First, the fixed effect-structure was determined. The initial mixed-effects model contained all experimental variables with all their interactions (up to 3-way) and control variables without interactions as fixed effects, and intercepts of subjects and items were included as random effects. Next, this model was reduced step-by-step. Significance of the interactions was evaluated by likelihood-ratio test with a 0.05 alpha-level for model-selection, where the model was compared to a similar model without the interaction of interest. In a second step, the reduced model was extended with the slopes and interactions of the fixed effects in the random effect structure. If the model with all random effects did not converge, a forward "best path" algorithm was used to evaluate which random slope (and interaction) should be included. P-values were derived from a likelihood-ratio test with a 0.4 alpha-level for model-selection (Barr, Levy, Scheppers & Tily, 2012). At the end of this second step, we arrived at a model with fixed effects from the first step and with an extended random effect structure. In third step, the significance of the fixed effects of this model was once again evaluated with model comparison by likelihood-ratio tests. The model was compared to a similar model without the fixed effect of interest. If any interaction between fixed-effects was no longer significant, it was removed from the model. The mixed-effects model was run with a maximum of 1000 iterations.

3.6.2.2 Additional analysis

We found in our main analysis that response times decreased with increasing final word duration. In order to interpret the effect of the final word durations we conducted an additional analysis. We evaluated whether final word durations affected the response time because the questions contained a turn-yielding cue. If such cues in questions with shorter final words are closer to question-ends, differences in response times could arise as a result of the position of the cue relative to turn-end (Figure 9).

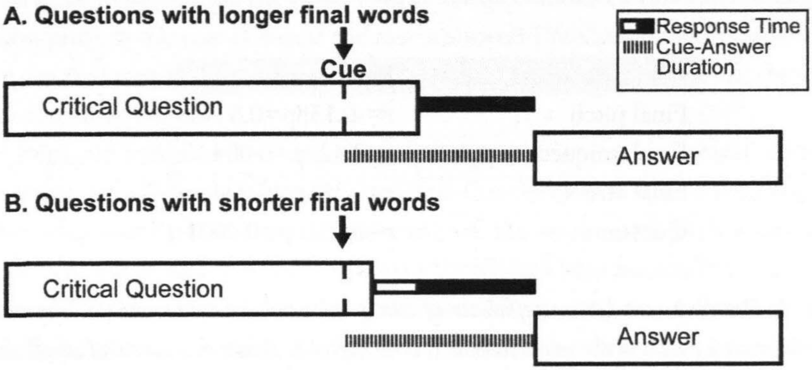


Figure 9. *A hypothetical model of the response times differences after shorter and longer questions. The model represents speech production when it starts as a reaction on the occurrence of a cue without a difference in temporal preparation. Duration of the critical question (left, white boxes) is longer in A) than in B). The cue (Cue) is closer to the end of the question in B) than in A). Response initiation is indicated by arrows. The duration between the cue and the beginning of the answer (striped line) is the same in A) and in B) which results in a difference in the duration between the end of the question and the answer (black bar, difference: non-filled).*

In contrast, if turn-yielding cues cannot explain alone the increasing response times at shorter final words, we can conclude that response times were influenced by a difference in the level of preparation for speaking at turns with different final word durations.

To test the effect of turn-yielding cues, we identified features in each question which could serve as potential cues for response execution. These cues were the following (1) the beginning of the questions, (2) the beginning of the final words, (3) the beginning of the vowel in the final stressed syllable, (4) the peak of the final rise in pitch, (5) the uniqueness point of the final word. The beginning of last words was measured from the ending of the preceding word. We calculated the duration between these possible cues and the question ends.

First, we tested if there was a correlation between final word duration and the duration from the cue to the question end. Cues which are closer to the turn end in shorter final words, are potential candidates for explaining the effect of final word durations on the response times in the main analysis. The duration of the questions ($r=0.851$, $p<0.001$) and the duration between the uniqueness point ($r=0.62$, $p=0.004$) and the question end showed significant correlation with the final word duration (Table 3).

	Final word
Final pitch	r=-0.13, p=0.6
Final uniqueness point	r=0.62, p=0.004
Final stress	r=0.38, p=0.094
Question	r=0.851, p<0.0001

Table 3. Pearson's correlation coefficients of the final word durations and the durations between possible cues and the question ends (N=20,df=18, Bonferroni corrected significance level: 0.0125). Final word: duration of the final word in the critical questions; final pitch: duration between the final pitch and the end of the critical questions; final uniqueness point: duration between the uniqueness point of the final word and the end of the critical questions; final stress: duration between the final stressed syllable and the end of the critical questions; Question: duration of the critical questions.

Then, we evaluated if we can still find an effect of final word durations when response times are calculated from these possible cues (uniqueness point, beginning of the final words and beginning of the questions). For this, a new response variable, Cue-Answer was calculated. This was the duration between the possible cue and the beginning of the answer (Figure 10).

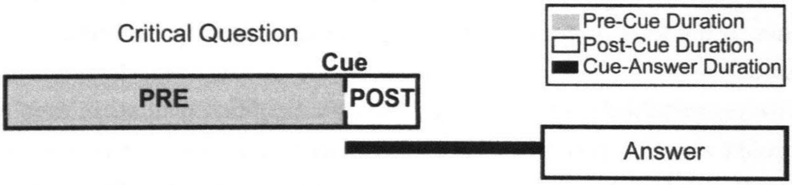


Figure 10. *Illustration of the Pre-, Post-Cue and Cue-Answer duration. Pre-Cue duration is measured between the start of the critical question and the cue (left box, light gray), Post-Cue duration is measured between the cue and the end of the critical question (left box, non-filled) and Cue-Answer duration is measured between the cue and the participant's answer (black bar).*

If the duration of Cue-Answer is influenced by the duration between the cue and the rest of the question (Post-Cue duration), this effect cannot be explained purely by reaction to cues. In contrast, if a response is initiated purely as a reaction to a cue, the duration of the rest of the question will not have any effect on the Cue-Answer duration. We also included into our analysis the duration between the beginning of the question and the cue (Pre-Cue duration) as a fixed effect (Figure 10). If participants react to a cue to initiate the response, temporal expectations can also arise based on the duration prior to the cue. If this is the case a shorter Pre-Cue duration leads to longer Cue-Answer duration. However, this effect would be independent from the duration of the turn after the cue.

Separate mixed-effects models were run with the Cue-Answer duration as independent variable and with the Pre- and Post-Cue duration as fixed effects respectively for the uniqueness point and the beginning of the final word. Before applying the regression model, it was checked if there is a correlation between the Pre- and Post-Cue durations. There were no significant correlations (uniqueness point: Pearson's r=0.028, p=0.907, N=20, df=18; beginning of final word: Pearson's r=0.19, p=0.418, N=20, df=18).

When the beginning of the question was tested as the potential cue, it was not possible to define the Pre-Cue duration. We followed the statistical modelling procedure described in Statistical Analysis. Table 4 summarizes the results.

Cue	Pre-Cue duration			Post-Cue duration		
	β	$\chi2(1)$	p	β	$\chi2(1)$	p
UP	-0.28	10.5	<0.01	0.48	11.76	<0.001
FW	-0.07	0.59	0.444	0.49	22.15	<0.0001
Q	n.a.	n.a.	n.a.	0.65	34.47	<0.0001

Table 4. *Beta-coefficient, chi-square and p-value of the dependent variables in the mixed-effects models of the uniqueness point of the final word (UP), the beginning of the final word (FW) and the beginning of the question (Q). Chi-square and p values were obtained by model-comparison (see Statistical Analysis) (Bonferrini-corrected significance level: 0.016).*

Table 4 shows that the Pre-Cue duration significantly affected the Cue-Answer duration when the cue was the uniqueness point. As the Pre-Cue duration increased, the Cue-Answer duration decreased. When the cue was the beginning of the last word, the Pre-Cue duration did not show a significant effect on the Cue-Answer durations. The Post-Cue duration showed a significant effect on the Cue-Answer duration in all cases. When the Post-Cue duration was longer, the Cue-Answer duration was also longer. The effect of the Post-Cue duration shows that the effect of final words on the response times cannot be explained only by the initiation of responses relative to the beginning of the question, to the uniqueness point or to the beginning of the final word.

The results show that possible cues we identified in our stimuli cannot alone explain the effect of the final word duration on the response times. This suggests that response times decreased at shorter last words because of the different level of preparation for the turn-ends. Trillenberg et al. has shown in interval-timing tasks that the level of response preparation is changing with the expectancy (the momentary probability) of the occurrence of the response signal (see also Niemi & Näätänen, 1981). Therefore, we were interested how such expectancy developed from the beginning of the last words in our experiment. Figure 11 shows the frequency of the duration of the final-words in 100 ms time-bins.

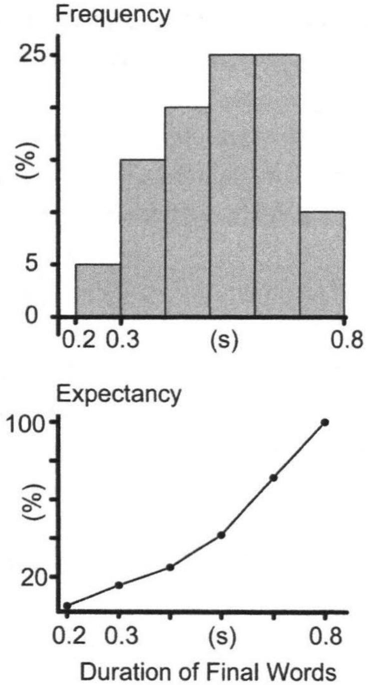


Figure 11. *The upper panel shows the frequency of the final word-durations in 100 ms bins in our experimental stimuli. The lower panel shows the momentary probability of the ending of final words (expectancy) in the next 100 ms based on the frequencies in the upper panel.*

Based on these frequency distributions, we calculated the probability of a turn-end falling into a time-bin given that the turn-end did not occur in an earlier time-bin. This reflects roughly how participant's expectations change about the occurrence of a turn-end with time. Figure 11 shows that the probability of the end of the last word is proportional to its duration. The higher level of expectancy of the word end can lead to higher level of preparation at longer final words which facilitates response times.

CHAPTER 4

4 EARLY ANTICIPATION LIES BEHIND THE SPEED OF RESPONSE IN CONVERSATION

Based on:
Magyari L., Bastiaansen, M.C.M., De Ruiter, J.P., & Levinson, S.C. (2014).
Early anticipation lies behind the speed of response in conversation.
Journal of Cognitive Neuroscience, 26(11), 2530-2539.

4.1 Abstract

Response times in conversation, with average gaps of 200ms and often less, beat standard reaction times, despite the complexity of response and the lag in speech production (600ms or more). This can only be achieved by anticipation of timing and content of turns in conversation, about which little is known. Using EEG, and an experimental task with conversational stimuli, we show that estimation of turn-durations are based on anticipating the way the turn would be completed. We found a neuronal correlate of turn-end anticipation localized in ACC and IPL, namely a beta-frequency desynchronization as early as 1250 ms, before the end of the turn. We suggest that anticipation of the other's utterance leads to accurately timed transitions in every day conversations.

4.2 Introduction

The primary ecology for language use and for the acquisition of language by children is the give and take of conversation. This conversational setting is characterized by rapid turn-taking, mostly with minimal gaps (under 200ms) between one speaker and the next (Stivers et al., 2009). Two additional properties make this coordination rather remarkable:

(1) a conversational turn is of no fixed length, adapting to the open-ended or generative character of natural language syntax (Sacks, Schegloff & Jefferson, 1974);

(2) the language production system is quite slow, even a single word requiring 600ms from conception to articulatory output (Levelt, 1989; Indefrey & Levelt, 2004), and multi-word utterances considerably longer (see e.g. Scnurr, Costa & Caramazza, 2006; Jescheniak, Schriefers & Hantsch, 2003).

If we put these facts together, it is clear that a would-be speaker must begin the production of his or her turn half a second or more before the other speaker has stopped speaking, and so must predict the end of the incoming turn even though it is of no fixed length.

There have been various proposals about how this remarkable coordination might be achieved. Some authors have suggested that there are turn-ending signals (analogous to the "over and out" on a two-way half-duplex radio), either in prosody (Local, Kelly & Wells, 1986; Local, Wells & Sebba, 1985, Cutler & Pearson, 1986; Beattie, Cutler & Pearson, 1982; Schegloff, 1996) or gaze (Kendon, 1967), but recent work does not support this for intonation (De Ruiter, Mitterer & Enfield, 2006) or gaze (Rossano, Brown & Levinson, 2009). Others have suggested that a composite bundle of turn-end features might be involved (Duncan, 1974). But all these suggestions run into problem (b) above, for the latency in the production system renders these signals too late to play a decisive role. Another suggestion is that turn-taking can be modelled by coupled oscillators (Wilson, 2005) on the basis of the speaker's rate of syllable production, in a manner similar to emergent coordination in e.g. fire-fly synchronization (Camazine et al., 2001). This suggestion runs into problem (a) above, that turns are not fixed in size, but have very varying durations. In addition, recent work shows that underlying even simple human synchronization there is a much more complex co-representation of joint coordination (Sebanz, Bekkering & Knoblich, 2006).

Thus, although we have a good grasp of the descriptive properties of the turn-taking system in conversation (Sacks, Schegloff & Jefferson, 1974), and evidence suggesting universal tendencies to minimize overlaps and gaps (Stivers et al., 2009), we do not understand the cognitive processes that make possible this virtuoso coordination which we all practice on the order of 1200 times a day (extrapolated from:Mehl, Vazire & Ramirez-Esparza, Slachter & Pennebaker, 2007).

The aim of the present study was to gain insight into the cognitive processes of the listener engaged in anticipating the ending of the incoming turn. We used the EEG signal of participants engaged in this task to explore the temporal dynamics of turn end anticipation – how far from the end of the turn does the listener move from a passive comprehension mode into a more active mode ready for the production of speech or action?

The current study builds especially on an earlier study (De Ruiter, Mitterer & Enfield, 2006) which experimentally assessed the relative contribution of intonation and lexico-syntactic content to turn-end prediction using turns extracted from natural conversation. Participants listened to each of these out of context and tried to press a key exactly at the ending of the turn. In the different experimental conditions, participants listened to (a) the original recording of a turn, (b) a version with intonational contour removed or (c) a version with no recognizable words but with intact intonation. When participants listened to the original recordings (a), they were able to press the key with an accuracy that paralleled turn-transitions in natural conversation, suggesting relatively little influence of pragmatic and context effects. Accuracy of the timing of key-presses did not change significantly when the intonation was filtered out. In contrast, when the words were rendered incomprehensible but the intonation was intact, the accuracy was greatly reduced. The authors concluded that people rely mainly on lexical and syntactic information for anticipating turn-ends.

How might lexical and syntactic information play a decisive role in predicting turn endings? While prosodic cues are assumed to appear just before the turn ends and to give only binary information to listeners whether a turn is ending soon or not yet, anticipated syntactic and lexical information is a good candidate for giving more fine-grained temporal information much earlier about when the turn is going to end. As a sentence unfolds the probabilities of continuations in different directions become ever narrower, a property exploited in nearly all modern machine processing of natural language (Bates, 1995; Manning & Schütze, 1999). Electrophysiological and eye-tracking studies have revealed that predictions are made during language comprehension at many different linguistic levels (Kamide, Altmann & Haywood, 2003; Wicha, Moreno, Kutas, 2004; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005; Altmann & Kamide (2007); DeLong, Urbach & Kutas, 2005; DeLong, Urbach, Groppe & Kutas, 2011). Listeners, as they process incoming turns, come to a point where they can actually predict the very next words (DeLong, Urbach & Kutas, 2005; DeLong, Urbach, Groppe & Kutas, 2011). Also, turns whose end-points can be more accurately predicted allow the prediction of the final words (Magyari & De Ruiter, 2012).

It is clear that listeners can predict the end of a turn before it ends. But it is unclear how early they sense the imminence of ending, and thus switch from a purely passive comprehending role into a more active role ready for speech or next action. These internal processes are not easy to get at through behavioural measures.

To explore the internal temporal dynamics, we used turns extracted from recordings of natural conversations as in the study (De Ruiter, Mitterer & Enfield, 2006) earlier described. A prior off-line gating task (see Methods), where participants had to complete actual turns cut short, was used to categorize turns as having either predictable (PRED) or unpredictable (UNPRED) final words during the last 600 ms before the turn end (Fig.1). For the main task, participants were asked to listen to the full turns in both conditions and try to press the key exactly at the end of the turn. We expected key presses to be more accurate for predictable turn ends. To reveal the temporal dynamics of turn-end anticipation, we measured the EEG of the participants while they were performing the experimental task. We expected to find anticipatory neural activity for predictable turn ends, not for unpredictable turn ends, appearing at least 600 ms before the turn end. We focused on the dynamics of EEG oscillations, as oscillatory dynamics in the alpha and beta frequency ranges have been clearly associated with both motor and non-motor anticipation in earlier research (Jasper & Penfield, 1949; Pfurtscheller & Aranibar, 1977; Pfurtscheller & Lopes da Silva, 1999; Bastiaansen & Brunia, 2001). Beta power and coherences changes have also been suggested to be related to syntactic and semantic processing (Bastiaansen, Magyari & Hagoort, 2010; Weiss et al., 2005; Wang, Zhu & Bastiaansen, 2012) and to reflect a close relationship between language comprehension and motor functions (Weiss & Mueller, 2012). We thus had two dependent measures, the timing of key-presses and the time-frequency analysis of EEG power changes.

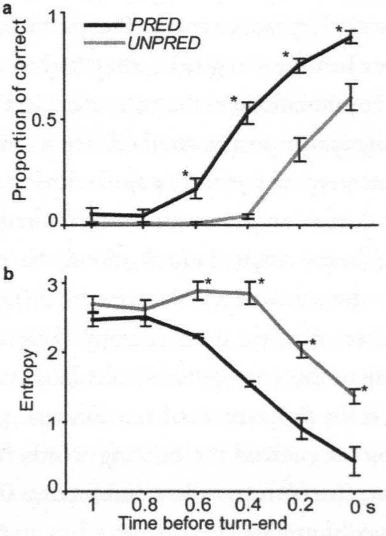


*Figure 1. Averaged results of the gating study for turns selected into the PRED and UNPRED conditions. The x axis shows how many seconds before the end of the turn the recording was cut off. Error bars indicate the standard error, * indicate significant differences between conditions. a) Proportion of correct answers averaged across turns of the two conditions at each gating points. b) Entropy of the answers averaged across turns of the two conditions at each gating points.*

4.3 Methods

4.3.1 Participants

Twenty-six participants (mean age 25 (range 19-39), 7 men, 15 women) gave informed consent and were paid for their participation in the EEG experiment. All were right-handed, native speakers of Dutch with no history of neurological or language disorders. None of them took part in the pretest of the stimuli material. Data from four participants were discarded due to excessive blinking, left-handedness or because of strikingly different key-press results that suggested that the participant did not follow the instructions.

4.3.2 Pretest of stimuli

The selection of the stimuli required a pretest using a gating paradigm. 48 participants from the subject-pool of the Max Planck Institute for Psycholinguistics participated in this study. None of them participated in the EEG study. Turns were used from Dutch, telephone-like conversations. The audio recordings of the conversations were made for another experiment (De Ruiter, Mitterer & Enfield, 2006). The recordings were made in two soundproof cabins in order to separate the channels carrying the recordings of the two speakers. For the pretest, the audio recordings of 108 turns were selected. These turns were 2.25 – 10 s long, were not followed by a laugh or breath, and were not interrupted by interjections from the other speaker. Each turn was cut 200, 400, 600, 800 and 1000 ms before the end. Each version of a turn (5 shorter and a full version) was assigned to different experimental lists. 8 participants per list performed in the experiment. Each list started with 12 practice turns. The participants were asked to listen to each segment once. After hearing a segment, they had to type on a computer keyboard their guess about the continuation of the turn starting from the last word that they heard. For further information on the method see a similar gating study, in Magyari & De Ruiter (2012). The answers were evaluated with regards to two aspects. First, each answer was coded as correct or incorrect, where an answer was correct if it exactly matched the words used in the original uncut stimuli. Second, it was also coded as to whether the answers to the same segment given by different participants were the same or different. Based on this, we used entropy (Shannon, 1948) to measure the variety of the answers. Shannon entropy was calculated using this formula: $\text{entropy} = - \Sigma \text{ pi log2 (pi)}$ where pi is the proportion of one kind of guess among the eight for each gating period (8 subjects guessed the missing words from each gating). If guesses are similar to each other, the entropy is low (minimum: 0), if the answers are different the entropy is high (maximum: 3).

4.3.3 Stimulus material

Based on the results of the gating study, 30 turns with the highest proportion of correct answers (mean = 0.404, (averaged across gating points)) were selected into the PRED condition of the experiment. These turns had also a low entropy across all gating points (mean = 1.688, averaged across gating points). Later, another 30 turns with a low proportion of correct answers (mean = 0.169) and with high entropy (mean = 2.415) were added to the UNPRED condition (differences in proportion of correct answers: t58=8.177, p < 0.001; differences in averaged entropy: t58=-6.899, p<0.001). The entropy and proportion of correct answers was different between the two conditions from the 600 ms gating point prior to the turn-end (t58 = 3.517, p = 0.001 (proportion of correct), t58 = -5.028, p < 0.001 (entropy) (Fig.1). Syllables were on average 178 ms, words 235 ms long. There was no significant difference in the duration of the turns in the two conditions (mean(PRED) = 4.25 s, mean(UNPRED) = 3.84 s, t58 = 1.015, p = 0.314).
An example from the PRED condition:
"Eh ik woon in een huis met vier vrouwen en nog een andere man" (Dutch)
('Eh I live in the same house with four women and with another man.' (Translation))
An example from the UNPRED condition:
"Oe en toen was ze weer eh s solo in eh in het noorden" (Dutch)
('Uh and then, she was again eh alone in eh in the north. (Translation))

4.3.4 Experiment and procedure

On the basis described above, 30 turns were selected into the PRED and 30 turns were selected into the UNPRED conditions. There were 22 other items that were selected originally for a third condition and 18 turns for practice. Data from these trials were not used for further analysis. Four experimental lists were created with different orders of the experimental trials. The practice trials were always at the beginning of each list, in the same order. Instructions and experimental task were similar to the instructions and task in De Ruiter et al.'s key-press experiment (De Ruiter, Mitterer & Enfield, 2006). Instructions appeared on the computer screen and contained the following (in Dutch): "The aim is that you should press the button PRECISELY at the moment the speaker finishes his turn. This means that you must try to predict the end of the fragment. You should not wait until the fragment has finished and then press the button." Participants were also instructed to avoid blinks and movements other than the key press during a trial. When participants pressed a green button the next trial started and a red button measured the responses. When an experimental trial started a fixation cross appeared on the screen, 1500 ms after which the audio fragment was played. A fixation cross was present until 2000 ms after the fragment finished or until the red button was pressed with the right hand. A blank screen was presented for minimum 1500 ms after the fixation cross indicating that the participant was allowed to blink. When the participants pressed the red button the audio stimuli stopped. When the black screen changed, a screen appeared with the instruction: "Press the green button!". Then the participants were free to start with the next trial. After the first half of the trials there was a break. Then, the experimenter went into the room and checked the participant and the electrodes. The experiment continued after the experimenter pressed a button outside of the room.

Participants were tested in a sound-proof, electromagnetically shielded room. They were seated at a distance of approximately 60 cm from a computer screen mounted on a table, next to a key-box with green and red response keys. The visual and auditory stimuli were played by Presentation software (version 12.1.03.24.08). Key-presses and the EEG were both recorded.

4.3.5 EEG recordings

EEG was recorded from 61 active Ag/AgCI electrodes using an actiCap. 59 of the 61 electrodes were mounted in the cap with equidistant electrode montage referenced to the left mastoid. Two separate electrodes were placed at the left and the right mastoid outside of the cap. Blinks were monitored through an electrode on the intra-orbital ridge below left eye. Horizontal eye movements were monitored through two electrodes in the cap placed approximately at each outer canthus. The ground electrode was placed on the forehead. Electrode impedance was kept below 10 kΩ.EEG and EOG recordings were amplified through BrainAmp DC amplifiers. DC recording was applied with a lowpass-filter of 100 Hz. The recording was digitized online with a sampling frequency of 500 Hz and stored for offline analysis.

4.3.6 Data preprocessing

Segmentation and artifact rejection of the EEG data was performed with Brain Vision Analyzer (version 1.05.0005) software. The data was segmented in epochs of 5000 ms, -3000 ms before and 2000 ms after key-press. A baseline between -2000 ms and -1500 ms before the key-press was used for artifact rejection. Approximately 23% of the trials were rejected. The average number of trials was 22.5 in PRED and 23.8 in the UNPRED conditions.

4.3.7 Behavioral data

The temporal offset between the end of a turn and the key-presses was measured. The averaged response time indicates how accurately subjects could anticipate the turn-ends. The averaged time is positive when participants press the key too late, and it is negative when participants press the key before the turn-end.

4.3.8 Time-frequency analysis of power

Time-frequency representations (TFRs) of single trial data were computed by using the multitaper approach (Mitra & Pesaran, 1999) with FieldTrip software package (Oostenveld, Fries, Maris & Schoffelen, 2011). TFRs show the power of the different frequency ranges at multiple time-points. Multitaper was applied first in a wider frequency range, and then the multitaper parameters were optimized for the beta frequency range. The final time-frequency analysis was done between 6 and 31 Hz in 1.25 Hz step-size and time steps of 10 ms with 5 Hz frequency smoothing and 800 ms time-smoothing. A relative baseline was applied on the TFRs between -2000 ms and -1700 ms before key-press. As a result of this the power values were expressed as the relative increase or decrease compared to baseline.

4.3.9 Source reconstruction

To identify the sources in the beta band, we used a beamforming approach, Dynamic Imaging of Coherent Sources (Gross et al., 2001). We were interested in localizing power differences between the conditions at the beginning and in the middle of the trials. Therefore, we created trials in both conditions that contained data from the 2 to 1.5 s before key-press (preperiod) and from 1.2 to 0.7 s before the key-press (postperiod). Based on the results of the time-frequency analysis, frequency analysis was applied using the multitaper method based on discrete prolate spheroidal sequences (Slepian sequences) on the trials at 15 Hz with a frequency smoothing of +- 3 Hz. Electrodes were aligned to a volume conduction model that was made based on a template brain using the boundary element method (Oostenveld, Praamstra, Stegeman, van Oosterom, 2001). A common spatial filter was then computed at 15 Hz for the different conditions and the pre- and postperiod together. The spatial filter was projected to all trials. Power values were calculated on an equidistant template 3D grid with a 5 mm resolution. Trials were averaged in the pre- and postperiods of the different conditions and the relative differences between conditions were calculated using the following formula: (powerpostperiod-powerpreperiod)/powerpostperiod. Finally, the grand-averages were computed and interpolated on the template brain.

4.3.10 Statistical analysis of behavioral results

Statistical significance of the differences between conditions in response times was evaluated by PASW Statistics 18. Repeated measures ANOVAs were computed on the averaged response times of each subject. Subjects' averages were calculated for the two conditions and for the first and second half the experiment. The ANOVA had two factors: condition (PRED vs. UNPRED) and order (first vs. second half of the experiment).

4.3.11 Statistical analysis of EEG results

For evaluating the differences between conditions in the EEG, we used a cluster-based random permutation procedure (Maris, Oostenveld, 2007) that is implemented in FieldTrip. We used this statistical approach because it elegantly handles multiple comparison problems. First, for every data point (sensor-time-frequency point) a simple dependent-samples t-test was performed that gave uncorrected p values. All data points that did not exceed a preset p value (here 0.05) were zeroed. Clusters of adjacent non-zero data points were computed, and for each cluster, cluster-level test statistics were calculated by taking the sum of all t-statistics within that cluster. A null distribution was created by randomly assigning the subject averages to one of the two conditions 1000 times, and for each of these randomizations a cluster-level statistic was computed. Then the largest cluster-level statistics of each randomization were entered into the null distribution. The observed cluster-level statistic was compared against the null distribution and clusters falling under the 2.5% of the two sides of the difference-distribution were considered to be significant. The statistical test was carried out between 2000 ms before and until the key-press.

For the statistical analysis of the source reconstruction, one-sided dependent-sample t-statistics were used comparing the power values of the trial-averaged subject data of PRED and UNPRED conditions at each sourcepoint which fall in the 3D grid within the template brain. There were 15711 grid points inside the brain, and for each grid point there were 6 neighbours (except at points at the edges of the brain where neighboring locations fall outside of the brain). Then, as a way of clustering, for sourcepoints that reached significance (uncorrected, $p<0.05$, $df=21$), we examined whether all of their neighbouring points were also significant. Voxels that had only significant neighbours were accepted as showing an effect. For localizing the spatial coordinates of the significant areas, the t-values of the significant, clustered sourcepoints and zeros at all other points were interpolated to a template brain (Oostenveld, Praamstra, Stegeman, van Oosterom, 2001). We identified brain areas using a template atlas (Lancaster et al., 1997).

4.4 Results

4.4.1 Behavioral data

Participants pressed the key on average 70 ms before the end of the turn in the PRED condition, but for the UNPRED condition they pressed the key on average 139 ms after the turn-end (see Fig.2). Figure 2 shows that there is a long negative tail in the distribution of the key-presses relative to the turn-end. Note, however, that the very early responses (1000 ms) before the turn-end which might be considered premature occurred only in a small percentage of the cases (5.3%). Moreover, all responses occurred after turn onset, and so, even in the case of very early responses participants probably tried to predict the turn end. The experimental condition showed a significant effect (F = 35.388, p = 0), but not the order of the presentation of the stimuli (F = 1.867, p = 0.186) and there was no significant interaction between condition and the order of stimulus presentation (F=0.255, p = 0.619). Thus, as expected, those turns whose actual final words could be predicted in a prior gating study, proved more predictable in an online reaction time task.

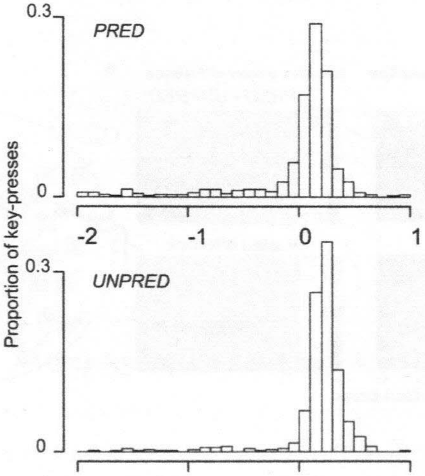


Figure 2. *Histogram of response times in the PRED and UNPRED condition. Response times were measured as the temporal offset between the key-presses and the end of turns. When the key was pressed before the turn ended, the response time is negative, when it was pressed after the turn-end, the response time is positive. The percentage of trials is shown on the y axis, and time in seconds before and after the key-press (key-press is at 0) is shown at the x-axis. The bars show the percentage of trials that falls into a 100 ms time-bin. Most of*

the key-presses fall into the 100-200 ms bias-bin in the PRED and into the 200-300 ms bin in the UNPRED condition. (Outlier responses smaller than -2 s and larger than 1 s are not shown.)

4.4.2 EEG data

4.4.2.1 Time-frequency analysis of power changes

The EEG signal showed a significant (p = 0.033) difference between the two conditions in the lower beta frequency range (11 to 18.5 Hz), starting around 1800 ms and lasting all the way up until the key-press (Fig. 3). A larger power decrease can be observed in the PRED condition. This difference was most prominent over midfrontal areas (Fig. 4).

Interestingly, the time course of beta power showed a different pattern over motor vs. midfrontal areas for the two conditions (Fig. 5). While beta power decreases were small (PRED) or non-existing (UNPRED) over the motor cortex, over midfrontal areas a strong decrease was associated with the PRED condition, and a strong increase with the UNPRED condition.

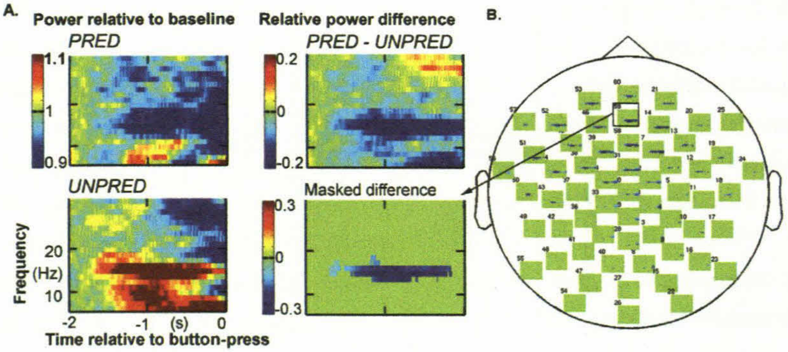


Figure 3. *Time-frequency representations of EEG power changes (TFRs). a) TFRs at electrode 59. The colorbars show the power values relative to baseline (from -2s until -1.7s). The first column shows the TFRs for each condition (PRED, UNPRED). The upper figure in the second column shows the relative power difference between conditions (PRED - UNPRED). The lower figure shows the significant power differences (MASKED). b) Schematic head with statistically masked TFRs at the corresponding electrode positions. The rectangle shows electrode 59.*

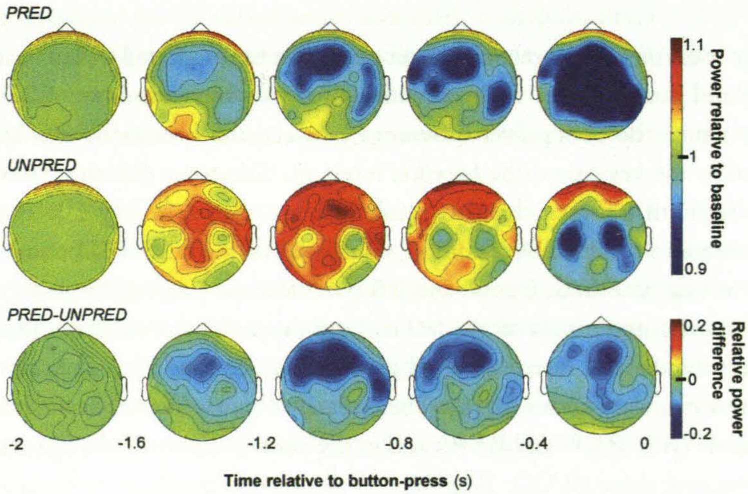


Figure 4. *Topographical distribution of beta band power (11-18.5 Hz) in subsequent bins of 400 ms. The upper and middle rows show beta power relative to baseline in the* PRED *and* UNPRED *conditions, respectively. The lower row shows the differences in power between the two conditions.*

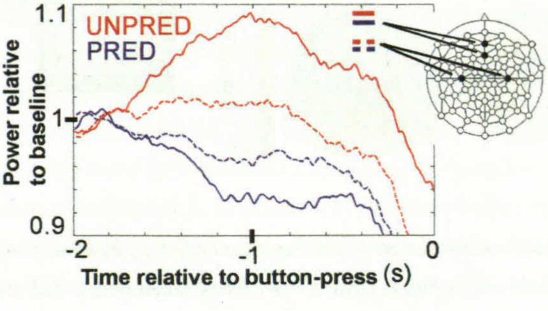


Figure 5. *Power values in the beta frequency range (11-18.5 Hz). Beta power is averaged across pairs of midfrontal (electrode 58,59, straight lines) and lateral central (electrode 37, 5, dotted lines) electrodes. Time is on the x-axis in s before the key-press (at 0), relative power values on the y-axis. Power is shown in red in the* UNPRED *condition and in blue in the* PRED *condition.*

4.4.2.2 Source reconstruction of the power changes

The source locations of the relative power changes were estimated with a beamformer technique and compared in both conditions for two time-windows: 1.2-0.7 s (the interval in which the beta power difference between the conditions was largest) vs. 2-1.5 s before the key-press (the baseline interval). The areas that show a difference in source strength between the two conditions are shown in Fig. 6B. The relative power decrease in the PRED condition, compared to the UNPRED condition, was estimated to originate from frontal and left parietal areas (Fig.6a). Frontally, a source is located in the anterior part of the left and right superior frontal gyrus that extends into the left middle and interior frontal gyrus (BA11 and BA47) and to the left and right anterior cingulate cortex (ACC). The parietal source is located in the left inferior parietal lobule (IPL, BA39 and BA40) and in the posterior part of the left middle and inferior temporal gyrus (BA37) (Fig.6b).

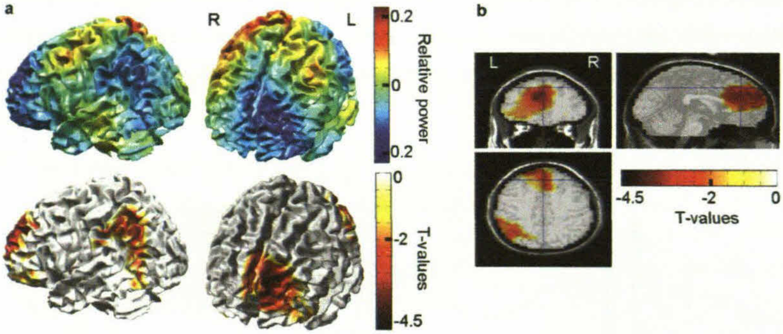


Figure 6. *Source reconstruction of the lower beta effect. a) Relative power changes (first row) and t-values of the source-points (second row) interpolated onto a 3D template brain surface. b) T-values of the source-points interpolated onto a template mri. Slices are shown at x=0, y=39, z=42 MNI coordinates.*

4.5 Discussion

Given the latency of the speech production process, if speakers are going to come in on time, they must begin the production process well before the end of the other's turn – and to time that, would-be speakers must predict the end-point of the incoming turn. As described, we used a prior gating task to sort turns into two kinds, relatively predictable or unpredictable, on the basis of whether their last words could be exactly predicted (Fig.1). In the main experiment, as expected, participants more accurately predicted the turns that were more easily completed in the gating study. The corresponding EEG signal showed that predictable turns, compared to less predictable turns, were accompanied by a power decrease in the beta band which is estimated to originate from left medial frontal, left superior frontal, left inferior parietal and left posterior temporal brain areas.

The behavioral measure, the timing of key-presses, is in line with the hypothesis that turn-end estimation matches the ability of participants to predict the actual last words of many turns starting from c. 600ms before turn-ending as shown in our prior gating study. It suggests that turn-end anticipation is built on predicting the actual forthcoming words. It would also allow just enough time for the production system to produce the first word of the response, given a 600ms production latency and an average turn gap of 200 ms. It would already rule out any role for late cues of turn-ending, such as turn-final prosodic cues.

However the EEG signal shows a much earlier anticipation of turn-ending. We found beta power differences during the anticipation of predictable (versus non-predictable) turn-ends already 1.8 s before the button-press. Allowing for the time-smoothing inherent to the time-frequency analysis (+- 400 ms), and the latencies of key-pressing (around +140 ms in the UNPRED condition), the observed differences in the EEG signal between conditions occurred on average at least 1250 ms before turn-ending. This means that people were anticipating the turn ends in the predictable condition at least more than 5 words before the turn-end on average (see average syllable and word duration in Methods, Stimulus material).

Turning to the interpretation of the EEG signals, it is well established that power decreases in the beta band can be observed during preparation for a movement above the sensorimotor areas (Jasper & Penfield, 1949; Pfurtscheller & Aranibar, 1977; Pfurtscheller & Lopes da Silva, 1999; Rektor, Sochůrková & Bočková, 2006; Alegre et al., 2006). Further, beta power decreases have been associated with the temporal predictability of stimulus occurance (Alegre et al., 2003; Alegre et al., 2006). Beta power and coherence changes have also been suggested to be related to syntactic and semantic processing (Bastiaansen, Magyari & Hagoort, 2010; Weiss et al., 2005; Wang et al. 2012; Weiss & Mueller, 2012).

The key-press results show that the difference in entropy correlated with turn-end predictions. Turns in the PRED condition had higher entropy which means there was a higher agreement about the continuations among participants during the pre-test of the turns. Higher agreement across participants presumably also correlates with higher confidence in the responses on the individual level in the PRED condition. Specially, if we also take into account that participants not only agreed less about the continuations in the UNPRED condition but there were also less correct guesses. More confident responses in the PRED condition could have resulted in differences in motor preparation. However, we found beta power decreases above the motor areas in both conditions as expected, but there were no differences across conditions above the motor areas. This indicates that a relative decrease in beta power in both conditions reflects motor preparation associated with the key-pressing, and that motor preparation processes are not different across the two conditions. Above frontal areas, however, there was a large beta power decrease during the predictable turns and a large increase during the unpredictable turns. These results show that neuronal correlates related to the anticipation of turn endings are distinct from those related to the anticipation of action.

The observed beta-band effects in the condition comparison might be thought to be a result of the differences in the predictability of the turn's content itself. However, empirical evidence shows that lexical predictability induces changes in gamma-band power, not beta-band power (Rommers, Dijkstra & Bastiaansen, 2013, Wang et al, 2012). Another possibility is that the turns in the unpredictable condition are less coherent which could lead to differences in the oscillatory activity. Bastiaansen et al.'s (2010) study shows that beta power increases throughout a correct sentence (correct and also coherent condition) compared to words presented in a random order (a less coherent condition). Therefore, if coherence plays a role in the observed EEG effect across conditions, we would expect to find higher beta power during predictable turns compared to unpredictable ones. However, instead of an increase we found beta decrease in the predictable condition. Therefore, the observed differences in the beta power across conditions most probably relate to the experimental manipulation, namely to turn-end predictions and not to differences in coherence.

In our study, we localized most of the beta power decrease to the left superior and middle frontal areas and the anterior cingulate cortex (ACC). This activation extended until the left middle frontal gyrus and the left inferior frontal gyrus (BA47). Another large locus of activation was found in the, left inferior parietal lobule (left IPL), left (posterior) middle and inferior temporal gyrus. During turn-end anticipation, the temporal estimation is based on the incoming linguistic information, which offers a different basis for prediction than other studies that have used time-estimation tasks (see e.g. Bastiaansen & Brunia, 2001). It is interesting therefore to try and delineate the functional brain network that subserves turn-end anticipation. The prefrontal cortex and the ACC are well known for being involved in anticipation and in time processing (Fuster, 2001; Bubic, Von Cramon & Schubotz, 2010; Aarts, Roelofs & van Turennout, 2008, Macar et al., 2002., Lewis & Miall, 2013), constituting a network of attentional control (MacLeod & MacDonald, 2000), verbal action planning (Hagoort, 2005;) and speech act comprehension (Egorova, Pulvermüller & Shtyrov, 2013). A left frontoparietal network involving the left intraparietal sulcus and left inferior premotor cortex has been suggested to be recruited particularly for directing attention toward a particular moment in time (Coull & Nobre, 1998). The IPL has been associated with the integration of incoming information into current syntactic and contextual frames (Lau & Poeppel, 2008). Brodman's area 47 has been involved in semantic unification, for example, in the integration of word meaning into the unfolding discourse context (Hagoort, 2005). The left posterior middle temporal gyrus and inferior temporal gyrus have been related to the activation and storage of lexical representations (Hagoort, 2005, Lau & Poeppel, 2008, Pulvermüller, 2005). Taking all these findings together, our present observation that the frontal, and left parietal, and temporal areas desynchronize in exactly the same frequency range as the motor cortical areas suggest a close coordination between brain areas subserving language comprehension processes, more general anticipatory behavior, and the motor network, during the execution of the experimental task.

The EEG data therefore show a clear, interpretable signal of early anticipation of turn-ending, based on the involvement of areas associated with syntactic, semantic and temporal processing. Although our experiment does not directly address the issue whether anticipation of turn-ends are based on prosodic or lexical/syntactic information, we selected our stimuli such that there was a difference between the predictability of the turn's lexical content between the different conditions from at least 600 ms before the turn-end. Prosodic cues are assumed to give information to listeners just before turn-ends on whether (1) the turn is ending soon or (2) it is not ending yet. In contrast, syntactic and lexical information are good candidates to give more fine-grained temporal information about when the turn is going to end. Based on our results, it seems likely that this information is available much earlier than turn-yielding prosodic cues. Syntax provides an architectural framework into which lexical material must slot, and as mentioned earlier, it provides ever narrowing completion probabilities as the incoming sentence is parsed (a process that seems to be reflected in our EEG measure towards the end of the turn), until a point where the precise final words can be anticipated (a point that seems to be reflected in our behavioral measure). Therefore, it is likely that turn-ends can be anticipated early based on lexical-syntactic information. These findings fit well into a Bayesian model of language processing, where the incoming linguistic material provides constant updating of expectations and narrowing likelihoods for alternative continuations (Chater & Manning, 2006; Christiansen & Chater, 2001; Friston, 2010). However, follow-up studies are needed to further narrow down the possible range of interpretations of the effects observed in this study.

This study has probed a little understood domain, namely how language is actually processed in its prime natural habitat, conversation. It suggests that, underlying the rapid turn-exchange system, anticipatory processing is required relatively early in the comprehension of a turn to achieve the apparently effortless coordination that is so commonly observed.

CHAPTER 5

5 MODULATION OF ALPHA SYNCHRONIZATION BEFORE SPEECH

5.1 Abstract

This study investigates whether late (post-lexical) stages of speech preparation can be timed to an expected moment of speaking and which oscillatory correlates are involved. Participants were asked to articulate non-words after an auditory stimulus (beep). In one block of the experiment, the beep was 2 second long, in the other block the beep was 4s long. Participants were slower to articulate the non-words after the end of beeps in the long beep condition. This probably reflects the less certain estimation of longer (4s) compared to shorter durations (2s). We found a strong synchronization in the alpha band observed at occipital areas during the later part of the first 2 seconds of the beeps compared to its beginning in both conditions. Alpha power (8-13 Hz) was higher at left temporal and occipital areas in the condition where the beeps were longer. We conclude that alpha synchronization probably reflects the inhibition of the visual areas in both conditions, presumably to provide more resources for processing of language and auditory information. The higher power values in the long beep condition might reflect stronger inhibition of the visual areas or inhibition of areas related to speech production.

5.2 Introduction

The interaction between understanding and producing speech plays a key role in the cognitive architecture which underlies fluent everyday conversations. Conversational partners frequently switch between speaker and listener roles, and so, between speech production and comprehension. The duration of turn-transitions, that is, the gaps and overlaps between the turns of the speakers are often really short. The most frequent turn-transition durations are between 0 and 200 milliseconds in conversations of several cultures and languages (De Ruiter, Mitterer & Enfield, 2006; Stivers et al., 2009). However, experimental studies have shown that the production of a word from conceptual preparation to articulation takes 600 milliseconds on average (Levelt, 1989; Indefrey & Levelt, 2004; Indefrey, 2011). Given the duration of the conversational gaps and duration of the speech production processes, speakers must often start speech production before the current turn they are listening to ends. Moreover, it is likely that speakers also anticipate and prepare for the moment when the previous turn ends (Sacks, Schegloff & Jefferson, 1974; Magyari & De Ruiter, 2012; Levinson, 2013). When the turn-end is anticipated, speakers deliver a response faster and more accurately timed to the end of the previous turn.

However, we know little about how speakers time speech preparation processes during conversations. Electro- or magnetoencephalography (EEG/MEG) which can trace neuro-cognitive processes with a high temporal resolution provide a useful means to reveal when preparation for speech production starts before overt articulation.

In contrast to speech comprehension, neuronal correlates of speaking are difficult to reveal because muscular activity causes a large effect on the EEG signal during speech production. To circumvent this problem, EEG studies have often focused on measuring the electrophysiological signal during speech preparation in the past twenty years (Van Turennout, Hagoort & Brown, 1998; Abdel Rahman & Sommer, 2003; Indefrey & Levelt, 2004; Indefrey, 2011). Nonetheless, the focus of studies has been not the neuronal correlates of speech preparation but the time course of the different stages (conceptual preparation, lexical, access, phonological processing and articulation) of the speech production process. These studies have examined differences in ERP components, such as the Lateralized Readiness Potential or the N200 in linguistic go/no-go tasks to measure the temporal distance between the stages of production. Most of them used a button-press paradigm without requiring their participants to produce any words (e.g., Van Turennout et al., 1998, Abdel Rahman & Sommer, 2003).

In recent years, the number of MEG/EEG studies using overt naming tasks has increased (Saarinen, Parviainen & Salmelin, 2006; Koester & Schiller, 2008; Tremblay, Schiller & Gracco, 2008; Strijkers, Costa & Thierry, 2009; Gehrig, Wibral, Arnold & Kell, 2012). Strijkers et al. (2009) measured the onset of lexical access in speech production by comparing event-related potential (ERP) differences elicited by word frequency and cognate status. In this experiment, bilingual speakers were naming pictures and their ERP was measured up to 550 ms after picture onset. Koester and Schiller (2008) studied morphological processing in a priming study of overt naming. They compared event-related potentials just after picture presentation up until 700 ms. Following the earlier research tradition, these studies focused on the differences in ERP waves to gain information about the timing of a certain production stage in production. Only a few EEG/MEG studies have paid attention to the neuronal correlates of attentional and motor-preparation preceding a speech production task compared to non-speech. These have revealed local functional networks of speech preparation by the analysis of oscillatory dynamics in the brain activity. Tremblay et al. (2007) compared oscillatory dynamics in EEG between speech production and key-board pressing. Alpha and beta desynchronization was observed before speech and before key-board presses. They have related alpha suppression to attentional processes and beta suppression to movement execution and selection. Salmelin, Schnitzler, Schmitz and Freund (2000) and Saarinen et al. (2006) have also related motor preparation for speech to suppression of (beta) 20 Hz oscillations in the motor cortex. In Saarinen et al.'s (2006) MEG study participants had to execute speech and non-speech facial movements as soon as a visual stimulus was presented. Beta suppression has been localized at the facial representation brain areas before (speech and non-speech) facial movements and it also stayed suppressed throughout the movement execution. Moreover, the desynchronization has been tied in time to the visual stimulus and not to the movement onset. Hence, beta suppression may reflect a preparatory process which is induced by the instruction or intention for movement. In a recent study, Gehrig et al. (2012) studied neuronal correlates of speech preparation comparing an overt and a covert reading task. They focused on the interval between a cue which gave information about the task (silent or loud reading) and the presentation of sentences which had to be read. In this interval, they show that preparation for overt speech production is associated with left-lateralized alpha and beta suppression in temporal brain areas and beta suppression in motor-related regions. They conclude that the speech production network is already set up even before the content of the production is known. The observed reduction in alpha and beta power changes indicated the preparation for speech already 350 ms after the instruction for speech (cue). Besides beta desynchronization, others have also found alpha suppression before speech. Rommers, Meyer, Piai & Huettig (2013) have presented uncompleted sentences followed by figures of objects (constituting the last word of the sentence) to participants. Participants named or passively viewed the presented objects. They have found alpha and beta suppression before naming compared to the passive condition. Piai, Roelofs & van der Meij (2012) have found beta power increase for categorically related stimuli relative to unrelated stimuli in a picture-word interference naming task. However, they relate the beta power change not to differences in motor preparation but to the semantic interference effect during lexical selection. In a similar paradigm, theta and alpha (4-10 Hz) band activity was found related to lexical activation and competition (Piai, Roelofs, Jensen, Schoffelen & Bonnefond, 2014).

To sum up, oscillatory correlates of speech preparation have been identified in the power changes of alpha, beta and theta bands. More precisely, power changes have been related to processes of lexical selection (Piai et al., 2012; Piai et al., 2014), to preparation for a change in the motor system (Engel & Fries, 2010; Gehrig et al., 2012) and to active disinhibition of task relevant brain areas before execution of speech movements (Tremblay et al., 2007; Jensen & Mazaheri, 2010).

However, whether speech preparation is timed to the expected moment when articulation can occur has rarely been investigated. During conversations, the time at which listeners know what to answer and the time at which they can start producing their turn can vary. It may happen that listeners already know what to answer but they wait with their response until the end of the other's turn. Given the fast and frequent turn-transition times observed in natural conversations, it is unlikely that speakers do not prepare for speaking before the end of the other's turn or before turn-yielding cues.

On the one hand, it is possible that speech preparation processes start already at the moment when listeners are able to prepare their answer and they sustain the preparatory activation until the moment when the articulation can start. In accordance with this, Bögels, Magyari and Levinson (in preparation) have found alpha suppression in a question-answer paradigm in the middle of questions when the answer for the questions was known. In this study, participants were listening to quiz questions which they had to answer as soon as the question ended. The questions were constructed so that the critical information which enabled the participant to answer was presented either in the middle or at the end of a question. Bögels et al. found lower alpha power quickly following the critical information even if the information was presented in the middle of questions. When the participants' task was only to remember the questions but not to answer them, no changes in the alpha power were observed. This result might suggest that there is an attentional orientation towards speech production immediately when the content of what is to be said is available independently of the time when articulation will take place.

On the other hand, it is also possible that listeners start the preparation for speech relative to when they expect the turn-end coming. In chapter 3, we have shown that speakers respond earlier when they can more certainly predict how long questions will be. This might suggest that the speech preparation process starts earlier by timing it to the turn end of the other speaker.

These two alternative time-courses of speech preparation might be reconcilable. The first stages of the speech preparation process, for example, conceptual preparation and lemma selection (Levelt, 1989; Indefrey & Levelt, 2004; Indefrey, 2011) might start immediately while post-lexical stages, such as form encoding and preparation for articulation are timed to the expected moment of articulation. It might be also possible that later stages of speech preparation only start as a reaction to a turn-yielding cue or to turn-ends, but the attentional orientation when this moment is expected speeds up the reaction.

Nonetheless, to gain more insight about speech preparation during conversations, first, it is crucial to know whether later stages of the speech preparation processes can be timed to upcoming expected cues and how such processes are reflected in changes of the oscillatory dynamics. With respect to preparation of non-speech movements, Alegre et al. (2003) asked participants to extend their wrist as soon as possible after hearing a tone. They found beta desynchronization which was effected by the timing of the estimated appearance of the instruction to move the wrist. In one condition, the tone was presented rhythmically always with the same interval. In the other condition, it was presented randomly. In the rhythmic condition, participants could predict when the tone was to be presented and when they would extend their wrist. In the random condition, the timing of the tone was more uncertain. During the sequence of the predictable rhythmic stimuli, beta desynchronization was found 1 second prior to the stimulus compared to the unpredictable random condition. They conclude that the pre-movement beta suppression is an indicator of motor preparation. In chapter 4, we have also found beta suppression preceding key-press responses when the timing of movement execution is more predictable based on incoming linguistic information. The beta desynchronization was interpreted as a non-motor correlate of timing based on a close coordination between brain areas subserving language comprehension processes, anticipatory behavior and the motor network.

The main question addressed in this paper is whether there are motor and non-motor components of later stages of speech preparation which are timed to the predicted moment of speaking and how these processes are reflected in the oscillatory dynamics. If there are oscillatory correlates which can show whether speech preparation is timed to an expected moment of articulation, these would allow for further research of the timing of speech in conversational turn-taking.

In order to see whether speech preparation can be timed, we varied temporal certainty in our experimental paradigm. We asked participants to pronounce non-words when a meaningless auditory stimulation (beep) ends. We used non-words in order to focus on the later stages (form encoding and articulation) of the speech production process. We presented participants with blocks of beeps lasting either 2 or 4 seconds. Following a short practice, participants should be able to predict when the beep will end and when they need to say the non-word in both conditions (short or long beeps). It is well established that longer pre-stimulus intervals when they are presented in blocks slow down reaction times. It has been argued that reaction times are longer because time-estimation is less certain at longer durations and therefore, temporal preparation for the response is less optimal. (Niemi & Näätänen, 1981; Müller-Gethmann, Ulrich & Rinkenauer, 2003). Therefore, we predicted that we will find differences in the speed of the vocal response times after the end of the auditory stimulus between the two conditions. If we find longer response times in the long beep condition, we can infer that some part of the speech preparation is timed because it is affected by temporal certainty.

In this case, we also predict differences in the oscillatory dynamics before the responses. If motor and non-motor components of the speech preparation process are timed, we predict power changes time-locked to stimulus onset during the short compared to the long beeps. More precisely, if speech preparation is timed, motor preparation starts earlier in the short beep condition. In this case, we predict larger beta desychronization in the short beep condition compared to the long beep condition, because motor preparation for speech production (e.g. Salmelin et al., 2000; Gehrig et al., 2012) has been related to beta suppression.

In chapter 4, we found beta power desynchronization preceding key-presses which was related to temporal certainty of the moment of movement execution. Based on this result, we also predict beta power desynchronization in the long beep condition preceding response onset if estimation of longer time intervals is less certain.

5.3 Methods

5.3.1 Participants

32 (8 male, 24 female) native speakers of Dutch participated in the experiment who were recruited from the participant pool of the MPI for Psycholinguistics. They were between 18 to 25 years old and did not have any hearing impairment. They gave informed consent before participating and received 8 euro's per hour for their participation. The data of 3 participants was excluded from the analysis because of the high number of experimental trials contaminated by artifacts in the EEG data.

5.3.2 Materials and design

We created four Dutch non-words: aben, emok, ienus, and oudong. We also created two beep sounds of 377 Hz using PRAAT. They differed only in duration, one was 2000 and the other 4000 ms long. The non-words were randomly presented to participants in a blocked design with two experimental lists. In list 1 participants first received a block of 80 trials (of which the first 10 were practice trials) with short beeps (2000 ms), followed by a block of 80 trials (of which 10 practice) with long beeps (4000 ms). The order of blocks was reversed for list 2.

5.3.3 Procedure

The experiment was preceded by another experimental task in which participants had to remember Dutch sentences. We do not present the results of this task here (see Bögels et al., in preparation). An experimental trial went as follows: Participants saw a fixation cross on the screen for 1000 ms, then the non-word stimulus was presented visually for 400 ms. When the non-word presentation ended, the beep started and lasted for 2000 (short block) or 4000 ms (long block). Participants were instructed to articulate the non-word immediately after the beep ended. The next trial was started by the experimenter with a button-press after the participant said the non-word.

5.3.4 EEG apparatus

EEG was recorded from 61 active Ag/AgCI electrodes using an actiCap (see e.g. Chapter 4). Of these, 59 electrodes were mounted in a cap (actiCap) with equidistant electrode montage referenced to the left mastoid. Two separate electrodes were placed at the left and the right mastoid outside of the cap. Blinks were monitored through a separate electrode placed below the left eye and one of the 59 electrodes in the cap. Horizontal eye movements were monitored through two separate electrodes placed at each outer canthus. The ground electrode was placed on the forehead. Electrode impedance was kept below 10 kΩ. EEG and EOG recordings were amplified through BrainAmp DC amplifiers. A lowpass-filter of 100 Hz was applied. The recording was digitized online with a sampling frequency of 500 Hz and stored for offline analysis.

5.3.5 Data-analysis

Responses which contained errors or hesitations in the production of the non-words were excluded from the data-analysis (1.5% of all responses).

For statistical evaluation of the behavioural results, we used a (linear) mixed-effects model with maximum likelihood estimation. The lme4 package (Bates, Maechler, Bolker & Walker, 2013) of R (version 3.0.2), an open-source language and environment for statistical computing was used for the statistical analysis (R Development Core Team, 2013). The significance of the fixed effects of the mixed-effects model was evaluated with model comparison by likelihood-ratio tests using the anova function.

We used the Fieldtrip Matlab toolbox (Oostenveld et al., 2011) for the preprocessing and statistical analysis of EEG data. Two different analyses were performed. First, we analyzed the dataset when experimental trials were time-locked to the beginning of the beep (0 ms). The EEG data was segmented into epochs of -0.6 until 2 s. For artifact rejection, the segments were low-pass filtered at 35 Hz and baseline corrected between -0.6 and -0.4 seconds. Then, trials containing eye-, muscle-movements, or other artifacts were identified. The trials containing artifacts were rejected from the data which was not low-pass filtered and not baseline corrected. Before time-frequency analysis, the trials were demeaned (by subtraction of the average of the entire trial) and linear trends were removed. Time-frequency analysis was performed between 0.25 and 1.75 s. The power was calculated between 4 and 30 Hz in steps of 1 Hz using a Hanning taper (Grandke, 1983) with a window of 500 ms for each trial, and then averaged. The first 500 ms (0.25 – 0.75 ms) was used as a baseline.

Second, we analyzed the dataset time-locked to the voice onset. Start of the voice was calculated automatically based on the spectrogram and amplitude of the sound recording using PRAAT. Trials which were responded to later than 1 second from the beep end were excluded. Then, EEG was segmented into epochs from -2.25 to 0.25 aligned to the voice onset (0). Artifacts rejection was done between -2 and -0.25 s as described above. Before time-frequency analysis, the trials were demeaned and linear trends were removed. Time-frequency analysis was performed between -2 and 0 s. The power was calculated between 4 and 30 Hz in steps of 1 Hz using a Hanning taper (Grandke, 1983) with a window of 500 ms for each trial and then averaged. The first 500 ms (-1.75 to -1.25) was used as a baseline.

To test for statistically significant differences between conditions, we used the cluster-based approach implemented in the Fieldtrip toolbox (Maris & Oostenveld, 2007). This robust method reduces the multiple-comparisons problem and controls the family-wise error rate across subjects in time and space. To examine differences between experimental conditions, paired t-tests are performed for each time-point, channel, and frequency (for time-frequency analyses) with a threshold of $p < .05$. Significant clusters in time, space and frequency are identified on the basis of proximity (neighbors) in all dimensions of the cluster. Cluster statistics are calculated by taking the sum of t-values in every cluster. To obtain a p-value for each cluster, a Monte Carlo method is used to compute cluster statistics with 1000 random partitions of the samples. The proportion of random partitions which results in larger cluster statistics than the observed one is the p-value. Two-tailed threshold of .05 was applied for the evaluation of significance.

5.4 Results

5.4.1 Behavioral results

Participants responded in 562 ms from the end of the beep sound on average (min= 202 ms, max=2650, n=4003). We used a linear mixed-effects model with the following fixed effects: condition (SHORT or LONG auditory stimulus), the order of blocks (whether the SHORT or LONG condition block was presented first) and the order of trials within a block (as a continuous variable). The interaction of condition and the order of the blocks were also included in the first model. Participant, type of non-word (aben, emok, ienus, or oudong), and the interaction of these two with condition were included as random effects. The interaction of the fixed effect of condition and block order was not significant (Chisq=0.7,df=1,p=0.4). Therefore, the interaction term was excluded from the model. The order of the trials within a block (t=-1.104, Chisq=1.20,df=1,p=0.27) and the order of the blocks (t=-1.028, Chisq=1.01, df=1, p=0.31) did not have an effect on the response latencies, but the condition did (t=-2.873,Chisq=7.21,df=1,p=0.007) (see Figure 1). In the SHORT condition (mean = 536 ms), participants answered faster than in the LONG condition (mean = 587 ms).

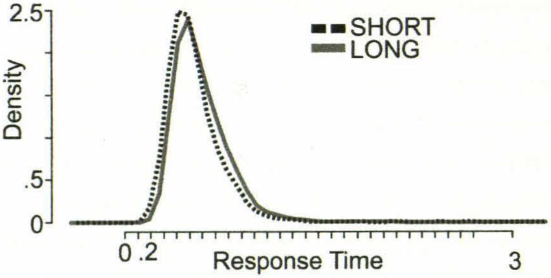


Figure 1. *Density plot of the vocal response times in the SHORT (black, dotted line) and in the LONG (gray line) condition. The end of the audio stimulus is at 0.*

5.4.2 EEG results

5.4.2.1 Time-locked to stimulus

Trials were time-locked to the beginning of the auditory stimulus (at 0 ms). After artifact-rejection, there was no difference across conditions in the number of trials (t=0.688,df=28,p=0.49). There were 53 trials per condition per participant on average with a minimum number of 31 trials. Trials were baselined between 0.25 and 0.75 s. Prior to applying a baseline, we computed cluster-based randomization statistics on the power values in the baseline window which showed no differences between conditions (first negative cluster: p= 1, first positive cluster: p= 0.6274). Cluster-based statistical analysis in the 0.25 – 1.75 window showed a significant difference between the SHORT and the LONG condition in the 8-13 Hz frequency band from 0.85 s until the end of the time-window (Figure 2). By visual inspection, we could see alpha band synchronization bilaterally above visual areas in both conditions compared to baseline (Figure 3). The significant cluster of the condition differences had an extended distribution encompassing almost all electrodes, but the strongest effect was located at left occipital and left temporal electrodes. There was a stronger synchronization in the alpha frequency band in the LONG compared to the SHORT condition. There was no desynchronization found in the beta band.

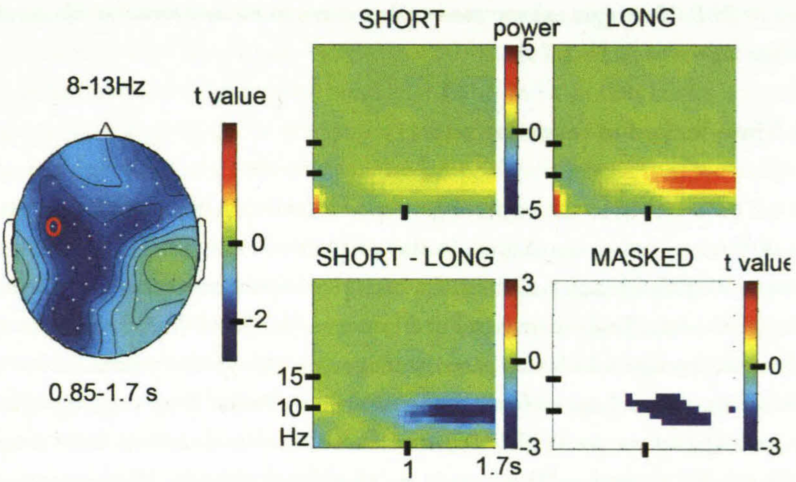


Figure 2. *Significant cluster of the differences between the SHORT and LONG conditions from 0.85 to 1.7 s in the 8-13 Hz band. The topography (left side) shows the electrodes which were part of the cluster (white). The T-values belonging to each electrode are color-coded. The red circle shows the electrodes to which the TFR's belong on the right side. TFR's are shown from 0.25 – 1.7 s between 4-30 Hz. Upper row, left: power values of SHORT condition; upper row, right: power in LONG condition; lower row, left: power differences between SHORT and LONG; lower row, right: t-values of the time and frequency points which are part of the significant cluster.*

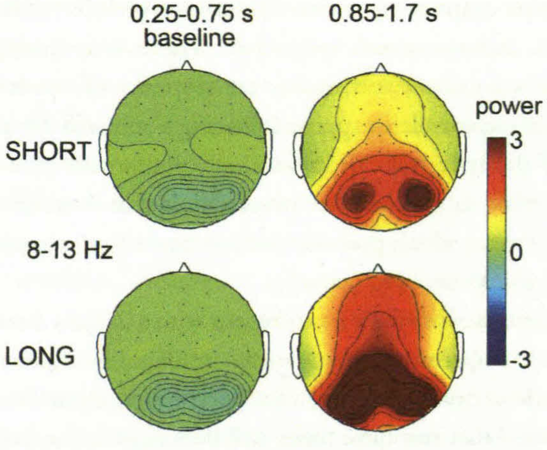


Figure 3. *Topographies of power values in SHORT (upper row) and LONG (lower row) condition in the 8-13 Hz frequency band. Left column: power values shown in the baseline*

period (0.25-0.75 s), right column: power values shown in the time-period of the significant condition differences (0.85-1.7 s).

5.4.2.2 Time-locked to response onset

Trials were time-locked to the beginning of speech (at 0 s). After artifact-rejection, there was no difference across conditions in the number of trials (t=0.018,df=28,p=0.99). There were 50 trials in each condition for each participant on average with a minimum number of 22 trials. Trials were baselined between -1.75 and -1.25 s. Prior to applying a baseline, we computed cluster-based statistics on the power values in the baseline window which showed no differences between conditions (first negative cluster: p= 1, first positive cluster: p= 0.2737). Cluster-based statistical analysis was computed in the -1.75 to -0.5 s window. There were no significant clusters (first negative cluster: p=0.2338, first positive cluster: p=1).

5.5 Discussion

The experiment was conducted to assess whether oscillatory correlates of speech preparation are timed to the predicted appearance of a response cue. During everyday conversations, the time when the speech production process can already start and the time when the articulation is executed vary. To advance the understanding of how speech production runs in real life, it is crucial to know whether speech production processes are timed and how these are reflected in the oscillatory dynamics. In the Introduction to this chapter, we reviewed studies which have shown beta and alpha desynchronization before speech or action. These were interpreted as neuronal correlates of the attentional and motor components of speech preparation. One possibility is that these correlates appear following a stimulus for a speaking task (even if the content of the speech is not known yet) or alternatively, at the moment when a speaker knows what to say. Another possibility is that they are correlates of motor or attentional processes which precede and are timed to the expected moment when execution of the articulation can occur.

Our experiment consisted of two blocks with a 2 and a 4 seconds long auditory stimulus (beep). Participants read non-words on the screen followed by a beep sound and they produced the non-words when the beep ended. Based on earlier button-press studies, we predicted that response times will be longer in the longer beep condition. We indeed found this effect, participants responded later in the long beep condition (Figure 1). It has been argued that reaction time is longer for longer time-intervals because longer intervals produce more temporal uncertainty which results in less optimal temporal preparation (Müller-Gethmann, Ulrich & Rinkenauer, 2003).

We also predicted alpha and beta desynchronization in the shorter beep condition compared to the other condition. We did not observe any desynchronization but we found synchronization in the alpha band (8-13 Hz) in both conditions instead. This synchronization was stronger in the longer beep condition as compared to the shorter one, already from 0.85s after the start of the beep. This suggests that the alpha power difference between conditions was tied to the expected time of articulation. In chapter 4, we have found beta power desynchronization preceding key-presses when the moment of movement execution was more predictable. Therefore, we also predicted differences in the beta frequency band prior to response onset in the two conditions. However, we found no differences when trials were time-locked to response onset.

Regarding the interpretation of the alpha power changes, alpha synchronization has been related to the functional inhibition of brain areas not engaged in a given task (Bastiaansen, Mazaheri & Jensen, 2012; Jensen & Mazaheri, 2010). In our study, a strong synchronization in the alpha band was observed at occipital areas during the auditory stimulus in both conditions. Therefore, it is likely alpha synchronization reflects the suppression of the visual areas during listening to the beep. Moreover, the alpha synchronization differences between conditions can also be interpreted in this framework. In the experiment, participants were asked to articulate the non-word stimulus as soon as possible after the beep ends. Therefore, the suppression of the visual system was probably needed in order to focus on the timing of the auditory stimulus. This is harder in the 4s condition, as it was shown in the behavioral data, hence, the inhibition was also stronger which is reflected in the higher alpha power.

However, it is also possible that the inhibition of the visual system occurred in order to focus on the demands of verbal working memory. The non-words were presented for 400 ms and disappeared from the screen when the beep started. Therefore, participants had to keep in mind the presented non-word. The stronger synchronization might reflect the larger demand of verbal working memory in the long beep condition because non-word stimuli had to be kept in mind longer. However, we only used four different non-words in the experiment all of which were easy to remember. Hence, the working memory demand of the task could be really low.

Moreover, if condition differences reflect a stronger inhibition of the visual areas, we would expect to find the strongest condition differences at occipital areas bilaterally. However, the condition differences were strongest at left electrodes, including left temporal areas. Recent models of word production describe a left lateralized network (Indefrey & Levelt, 2004; Indefrey, 2011). Papoutsi et al. (2009) has localized postlexical processes, phonetic encoding and the generation of the articulary code at bilateral but strongly left lateralized regions including mid and superior temporal and frontal regions, the premotor cortex and supplementary motor area. Therefore, it might also be possible that the stronger synchronization in the long beep condition reflects inhibition of areas related to the speech preparation processes. This means that participants inhibited their responses more strongly in the first 2 seconds of the long beep condition because the verbal response was longer delayed compared to the short beep condition.

To sum up, it is likely that alpha synchronization reflects inhibition of visual areas. However, the interpretation of why visual areas were suppressed and the interpretation of the condition differences require further research. However, we have observed a difference in the speed of response in the two conditions. Although we could not find an oscillatory effect which we could relate to earlier speech preparation in the short beep condition compared to the long beep condition (Gehrig et al., 2012; Salmelin et al., 2000; Saarinen et al., 2006), attentional orientation might account for the behavioural difference. Effects of temporal predictions are not restricted only to motor behavior. Temporal certainty can also modulate attention and perceptual processes (Nobre, Correa & Coull, 2007). This could facilitate the identification of cues (in this case, the end of beep) for which one must react.

We must be careful in the generalization of the results of this experiment to everyday language processing. Nonetheless, in terms of information processing during conversation, the results suggest that certainty in the temporal estimation of turn-durations could facilitate the speed of verbal responses. In conversation, however, turn-durations are not fixed; therefore, temporal estimation might be based on other (linguistic or contextual) information. The results might also suggest that speakers actively inhibit their responses when they already know what to respond but the other's turn has not ended yet.

CHAPTER 6

6 SUMMARY AND DISCUSSION

In this chapter, I summarize the results of the previous chapters and I discuss them in a broader perspective. I consider the limitations and the significance of the findings and their implications for further research.

6.1 Summary of the results

The aim of the series of experiments presented in this thesis was to examine whether temporal estimation based on the prediction of the turns' content affects the timing of conversational turn-taking. Timing is essential in our interactions with the external world, in the coordination of our own movements or in the coordination of our action with those of others (Sebantz, Bekkering, Knoblich, 2006). Social actions are implemented through speech and they are organized into an interactionally managed system of turn-taking (Schegloff, 2007, Levinson, 2013). The timing pattern of the turn-taking system suggests that conversational partners attempt to coordinate their turns not only in content but also in time (Sacks et al, 1974; De Ruiter et al, 2006; Stivers et al, 2009.). Hence, this thesis focused on whether speakers in conversations prepare for the moment when the other speaker' turn ends. More precisely, the key issue under investigation was whether turn durations are predicted based on the anticipated lexical and syntactic content of turns.

Chapter 2 examined this question by measuring whether listeners can anticipate the final words of conversational turns and whether these predictions correlate with turn-end predictions measured by button-press. We used the stimuli and button-press results from an earlier study (De Ruiter et al., 2006). In De Ruiter et al.'s study the participants' task was to listen to single turns taken from recordings of natural conversations and attempt to press a button precisely when the turns ended. In our study, we presented the initial fragments of these turns to participants and asked them to guess how the turns would continue. We have found a positive correlation between the accuracy of the button-press results of the earlier study and the accuracy of prediction of words and prediction of the number of words. We concluded that turn-end predictions are helped by the anticipated syntactic and lexical properties of the turn's content.

The study in chapter 3 investigated whether the predictability of the length of the last word of questions affects turn-transition times. Participants were asked to answer questions as soon as they ended. In the different experimental conditions, we manipulated whether it was possible to predict the length (i.e. the number of syllables) of the last word of the questions in advance, and whether a correct answer was clear without listening to the entire question. When participants knew the answer earlier, they answered with a shorter gap after the question's end. Response times also decreased when the last word of the questions were longer. This effect was in interaction with the predictability of the length. Participants answered questions with short final words more quickly when they could predict the length of the final word. When the last word of a question was long, predictability did not affect response times. We proposed that the level of response preparation increased with longer duration and with the predictability of the duration of the final words. However, when the last word of the question was long, participants had enough time to reorient their attention and response preparation. Hence, predictability only affected response latencies when the questions ended with a short word. We also concluded that participants perhaps prepared the answer before the last word when they knew the answer in advance.

Chapter 4 investigated the neuronal correlates associated with the predictions of turn-ends. We asked participants to listen to turns with highly predictable and unpredictable last words and to press a button when the turns ended. We found that button-presses occurred later when the last words of turns were unpredictable. In addition, we also measured the EEG of the participants. We found lower power in the beta frequency range for predictable turns. This effect occurred more than a second before the response. The differences between conditions were localized at frontal (including anterior cingulate cortex) and at left temporal and inferior parietal brain areas. We suggested that the brain areas show a local network involving both the timing and language processes which desynchronize in the same frequency band.

Chapter 5 examined when speakers start to prepare speech. The research question was whether speakers time the start of speech preparation relative to the predicted moment they have to start speaking. To investigate this, EEG was used in a delayed naming paradigm. Participants were shown written Dutch non-words which they were asked to produce when an auditory stimulus (beep) ended. In one block, the beep lasted 2 seconds; in the other block, the beep was 4 seconds long. Verbal responses were faster following the shorter (2s) beep. This suggests that uncertainty induced by the longer duration increased the speed of verbal responses (Niemi & Näätänen, 1981; Müller-Gethmann, Ulrich & Rinkenauer, 2003). We also observed alpha synchronization preceding the end of the beeps bilaterally at occipital areas. We suggested that alpha synchronization indicates the inhibition of the visual areas while listening to the beeps. We also found higher alpha power in the long beep condition which was strongest at electrodes above left occipital, parietal and temporal areas. We suggested that this reflects stronger suppression of the brain areas related to visual processing or to speech production.

To sum up, the behavioural results of the studies established that prediction of a turn's content (i.e. anticipation of the words, number of words or number of syllables of the last words of turns) affects the anticipation of the turn-end. When the turns' content can be anticipated with certainty, response latencies are faster. Response latencies are also faster when speakers know early what to respond to an incoming turn. Finally, the last two studies revealed oscillatory correlates related to the prediction of turn-ends (Chapter 4) and to attentional or inhibitory processes preceding speech (Chapter 5).

6.2 Discussion

This section discusses the interpretation of the experimental results and the limits of our understanding with regard to the research questions considered in the introduction. At the end, a model of timing of conversational turns will be presented.

6.2.1. The framework of the thesis and the research questions

This thesis started with the observation that turn-transition times in natural conversations are often short (around 200 ms). The introduction presented the central question of how speakers are able to respond so quickly. It was proposed that listeners can prepare their turn during a turn in progress, and can time their speech onset by anticipating the end of the current turn based on its predicted lexical and syntactic content.

It was argued that next speakers must prepare their turn during the current turn given the relatively long latencies of the speech production. This argument was based on the results of experimental research which has shown that much longer time (at least 600ms) is required to retrieve and code a word in preparation for articulation (Levelt, 1989:222) than the duration of the most frequent turn-transitions. The research questions of this dissertation thus focused on the anticipation of turn-ends, an issue not frequently studied with experimental methods. Nonetheless, the study presented in chapter 3 showed that speakers are quicker to respond when they know the answer for questions in advance. This confirmed that when possible, speech preparation begins during an incoming turn.

The starting hypothesis was that interactants are also able to predict turn-ends based on the predicted syntactic and lexical content of turns. Accordingly, the first research question of the dissertation was whether listeners can predict the lexical and syntactic content of conversational turns. Second, it was considered whether the predicted lexico-syntactic content enables listeners to anticipate turn-ends. Third, the neural correlates of turn-end predictions were examined in an EEG study that probed how early during a turn listeners can anticipate turn-endings. The final research question concerned the nature of the anticipatory mechanism that helps speakers time their turns closely to the end of the previous turn.

6.2.2. Research question 1: Are listeners able to predict the lexical and syntactic content of conversational turns?

The studies in chapter 2 and 4 showed that listeners are able to correctly guess both the final words and the number of words of conversational turns. Previous experimental studies have shown that listeners can make predictions during language comprehension, but these experiments often used controlled linguistic materials with strong contextual constraints. Here, we showed that unconstrained natural speech is also predictable to a certain degree.

We used a rough measure (listeners' predictions of the number of words) to infer their expectations about the syntactic structure of turns. It is possible that semantic or pragmatic constraints played an additional role in these predictions. Further research could reveal the relative contribution of these information sources.

6.2.3. Research question 2: Does the prediction of the content affect the anticipation of the turn-end?

The studies in chapter 2 and 4 showed that the prediction of turn-ends was more accurate when the content of turns was more predictable. These experiments used button-presses to measure the accuracy of turn-end anticipation. The study in chapter 3 showed that the predictability of word length also influenced verbal response times. These results suggest that the prediction of a turn's content leads to better prediction of the turn-end and to shorter turn-transition times.

Further research may explore how people estimate durations based on the length of the anticipated content, how the duration of the predicted content is represented and whether the speech-rate is taken into account in these estimations. So far, most models of language processing have not considered whether listeners predict the duration of speech, a notable exception being Pickering and Garrod's (2013) model. According to their model, interactants use a forward model to predict the other's and their own speech. The forward model also represents the latency of the preparation of the predicted speech. However, the authors do not elaborate how the predicted latencies could facilitate timing of turns during conversation.

Although the experiments of this thesis tapped into basic processes of anticipation of turn-ends, further research could investigate whether these processes also operate in real life situations. The experimental setup included instructions such as to answer as soon as the question ended or to press a button exactly at turn-ends. These instructions might explicitly orient participant's attention to response speed. We know little of the extent to which speakers are aware of the timing of their speech in natural conversation. The analysis of corpus data or experimental manipulations in a natural context could eliminate the effect of attention induced by the experimental instructions.

6.2.4. Research question 3: What are the neuronal correlates of turn-end anticipation? How early do expectations arise about the turn-end?

A neuronal correlate of turn-end anticipation was revealed in the oscillatory dynamics in chapter 4. Differences in the oscillatory dynamics occurred more than a second (ca. 1250 ms) before the turn-end. These findings support the idea that participants anticipate and predict turn-endings, rather than react to turn-final cues. The recorded oscilatory dynamics confirm that speakers can anticipate a turn-end prior to the presence of turn-yielding cues.

In addition to evaluating the time-course of turn-end anticipation, this study also used source-localization to implicate some responsible brain areas. EEG source-analysis can only provide limited information concerning the cortical regions involved in cognitive processes, and so further research is necessary to reveal the entire functional brain network involved in turn-end anticipation. Techniques with better spatial resolution (e.g. fMRI or MEG) could provide more detailed information concerning the involved brain regions.

6.2.5. Research question 4: What is the nature of the anticipatory mechanism which leads to tight turn-transitions?

The introduction (section 1.3.2) discussed how anticipation of turn-ends facilitates the timing of speech. One possible mechanism is that speakers start the speech preparation process relative to the estimated turn-end. Given that the speech preparation process takes ca. 600 ms, if listeners anticipate that the current turn will end in 1000 ms and can already conceptualize a response, then they will start the speech production process in 400 ms in order to produce their turn on time. This account assumes that the listeners (next speakers) are aware of the latency of their own speech preparation process. Pickering and Garrod (2013) propose that speakers calculate a forward model of their own speech, and this forward model also represents the latency of the speech preparation. In contrast with this, Levelt (1983) suggests that speakers have no access to the content of the speech preparation process, and that they can only monitor their covert speech (i.e. their not-yet-articulated speech). If this is the case, it is likely that speakers cannot control the timing of speech preparation or represent the duration of this process.

The study of chapter 5 could not confirm that speech preparation begins relative to turn-end. Although there was a difference between conditions in the alpha power, this difference could either be related to the suppression of visual areas, or to the inhibition of speech production areas. Nevertheless, verbal response times were affected by the duration of the experimental trials; reaction times were longer for longer time-intervals. This suggests that anticipation of when to speak affects verbal response times, but not by timing the speech preparation process relative to the turn-end.

Based on this study and on the findings of another recent experiment (Bögels, et al., submitted), a different anticipatory mechanism can be proposed. Speakers probably prepare a next turn as soon as they can predict the content of the current turn. However, listeners may have to delay the articulation of their responses so as not to overlap with the current speaker. They would then start articulation once they have enough evidence that the current turn is coming to an end. Articulation will take less time if the next turn has already been prepared. Moreover, articulation could also be faster if the speaker can anticipate when to start. Anticipation of the turn-end will facilitate the level of attentional and motor preparation for articulation onset. With regard to turn-taking, syntactic and lexical information could provide a base for temporal preparation of articulation.

Unexpectedly, the study in chapter 3 also showed that verbal response times are affected by the duration of questions independently of their content. The experiment showed that longer turn durations lead to faster response times. This is in contrast with the findings of chapter 5 in which verbal responses were faster following the condition with shorter experimental trials (2s beep).

The different effects of the longer durations of these two experiments are not surprising if we turn to studies of interval-timing. In these studies, temporal preparation is manipulated by varying the interval between a warning signal and the response signal. When the interval between the two signals (the so-called foreperiod) is constant throughout a block of experimental trials (as in the study in chapter 5), response times are usually longer for longer foreperiods. Researchers proposed that longer intervals lead to more temporal uncertainty resulting in less optimal temporal preparation for reaction (Müller-Gethmann, et al., 2003). When the foreperiod randomly varies from trial to trial (as in chapter 3), reaction times vary according to the probability of the stimulus occurrence. Participants become tuned to these probabilities and the level of their preparation varies accordingly (Niemi & Nätäänen, 1981; Nobre, et al., 2007; see also Müller-Gethmann, et al., 2003). In chapter 3, we proposed that, as the question continued, the probability of it ending increased, which led to increasing temporal preparation and faster verbal responses. We also note here that turn-durations vary in natural conversations as well. Further research could examine whether the conditional probabilities of turn duration in conversations also influence the next speaker's motor and perceptual preparation for articulation.

6.2.6. A model of timing of turns in conversation

The studies of this dissertation showed that listeners can predict the syntactic and lexical content of conversational turns. They also showed that the anticipation of the syntactic and lexical content helps interactants anticipate the end of current turns, facilitating faster turn-transitions. Based on the findings of this dissertation, the following model of timing of turns can be proposed:

Interactants can often predict the syntactic and lexical content of turns in natural conversations. This helps the next speaker achieve tight turn-transitions in two ways. First, predicting turn content enables the next speaker to conceptualize a response and start to prepare the next turn. The next speaker will start with speech preparation as soon as a response can be conceptualized. Articulation of this next turn can be delayed in order to avoid overlap with the current speaker. Second, the next speaker can also anticipate the end of an incoming turn based on the length of its predicted syntactic and lexical content. Turn-end anticipation might be also influenced by the overall distribution of turn-durations in conversations. Anticipating when a turn will end allows the next speaker to optimize motor and attentional preparation for articulation, which facilitates the speed of articulation.

The level of the speaker's preparation for articulation is a function of the predicted length of the incoming turn, and the certainty of that prediction. When the turn-end is predicted to be approaching, the level of preparation will be higher. When this prediction is uncertain, the preparation level will be lower. The studies presented in this dissertation have indeed shown that key-press and verbal response times varied according to participants' certainty in predicting the final words of the turns. Given the continuity of the incoming linguistic input, these predictions can be continuously updated and the preparation can be adjusted accordingly. Such an account is not incompatible with the turn-yielding cues account, as turn-final cues also increase the probability that the turn is ending. However, this account suggests that next speakers do not merely wait to encounter a turn-yielding cue and react to it, but prepare what they will say and when they will say it in advance.

It is important to note that temporal preparation is possible even when the response is not fully specified. Reaction time studies have shown that reaction times are faster if the timing of the response signal is predictable in advance but not the type of the response (e.g. which hand to move) (see Jepma et al., 2011). This observation aligns with our finding that speakers' early or late knowledge of the answer was independent of the effect of the predictability of the turn-ends (chapter 3). This also coincides with the findings of a study by Gehrig and colleagues (2012), which demonstrated that the speech production network can begin preparation before the actual content of the speech is available. Hence, this suggests that the prediction of turn-ends can facilitate turn-transition times even if speech preparation starts late.

Further research could examine whether turn-end anticipation is affected by attentional constraints and working memory load. Interval timing studies show that the motor behavior and the perceptual system are automatically attuned to conditional probabilities and to regularities of when events occur (Nobre, et al., 2007). These temporal expectations are unaffected by other task demands. However, when symbolic cues are used to indicate the duration of the interval preceding a response signal, preparation for the response decreases with increasing task demands, for example, with increasing working-memory load (Rohenkohl, et al., 2011; Capizzi et al., 2012). Therefore, it is possible that temporal preparation for turn-end based on the predicted content of turns might decrease by attentional demands and parallel processes, for example, by difficulties in speech preparation.

This thesis has explored the role of anticipation of the turns' content in the timing of turn-taking. The studies showed that speakers process anticipated information and adjust their response preparation and attention accordingly. Linguistic anticipation has typically been studied with well-formed, linguistic constructions. Here, we have shown that anticipation of the upcoming content is also possible using natural speech, and that this anticipation has an effect on the timing of responses. Consequently, linguistic anticipation not only allows for faster information processing in real life as has often been argued (Kutas et al., 2011; Garrod & Pickering, 2004), but is also crucial for timing in social interactions.

7 References

Aarts, E., Roelofs, A., & van Turennout, M. (2008). Anticipatory Activity in Anterior Cingulate Cortex Can Be Independent of Conflict and Error Likelihood. *Journal of Neuroscience*, *28*(18), 4671–4678. doi:10.1523/JNEUROSCI.4400-07.2008

Abdel Rahman, R., & Sommer, W. (2003). Does phonological encoding in speech production always follow the retreival of semantic knowledge? Electrophysiological evidence for paralell processing. *Cognitive Brain Research*, *16*(3), 372–382.

Alegre, M., Imirizaldu, L., Valencia, M., Iriarte, J., Arcocha, J., & Artieda, J. (2006). Alpha and beta changes in cortical oscillatory activity in a go/no go randomly-delayed-response choice reaction time paradigm. *Clinical Neurophysiology*, *117*(1), 16–25. doi:10.1016/j.clinph.2005.08.030

Alegre, M., Gurtubay, I.G., Labarga, A., Iriarte, J., Malanda, A., & Artieda, J.. (2003). Alpha and beta oscillatory changes during stimulus-induced movement paradigms: e¡ect of stimulus predictability. *Neuroreport*, *14*(3), 381–385.

Altmann, G.M.T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502–518.

Altmann, G.T.M., & Kamide, Y.. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Anderson, A.H., Bader, M., Gurman Bard, E., Boyle, E., Doherty, G., Garrod, S., … Weinert, R. (1991). The Hcrc Map Task Corpus. *Language and Speech*, *34*(4), 351–366.

Baayen, R.H. (2008). *Analyzing Linguistics Data: A practical introduction to statistics.* Cambridge: Cambridge University Press.

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society B, 364*, 1181-1182.

Barr, D.J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18–40.

Barr, D.J., Levy, R., Scheppers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bastiaansen, M.C.M., & Brunia, C.H.M. (2001). Anticipatory attention: an event-related desynchronization approach. *International Journal of Psychophysiology, 43*, 91–107.

Bastiaansen, M.C.M., Magyari, L., & Hagoort, P.. (2010). Syntactic Unification Operations Are Reflected In Oscillatory Dynamics during On-line Sentence Comprehension. *Journal of Cognitive Neuroscience, 22*(7), 1333–1347.

Bastiaansen, M.C.M., Mazaheri, A., & Jensen, O.. (2012). Beyond ERPs: Oscillatory neuronal dynamics. In Luck, S.J. & Kappenman, E.S. (Eds.), *The Oxford handbook of event-related potential components* (pp. 31–50). New York: Oxford University Press.

Bates, D.M., Maechler, M., Bolker, B., & Walker, S. (2013). *Lme4: Linear mixed-effects model using Eigen and S4*. Retrieved from http://CRAN.R-project.org/package=lme4

Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. Retrieved from http://CRAN.R-project.org/package=lme4

Bates, M. (1995). Models of natural language processing. *Proceedings of the National Academy of Sciences, 92*(22), 9977–9982.

Beattie, G., Cutler, A., & Pearson, M. (1982). Why is Mrs. Thatcher interrupted so often? *Nature, 300*, 744–747.

Blaauw, E. (1991). Phonetic characteristics of spontaneous and read-aloud speech. Presented at the ESCA Workshop on Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication, Barcelona.

Bock, K., Eberhard, K.M., & Cutting, J.C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language, 51*(2), 251-278.

Bögels, S., Magyari, L., & Levinson, S.C. (submitted). Neuronal correlates of speech preparation in an interactive quiz paradigm.

Bögels, S., & Torreira, F. (submitted). Intonational phrasing is necessary for turn-taking in spoken interaction.

Branigan, H.P., Pickering, M.J., & Cleland, A.A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*, B13–25.

Branigan, H.P., Cathpole, C.M., & Pickering, M.J. (2010). What makes dialogues easy to understand? *Language and Cognitive Processes, 26*, 1667–1686.

Bubic, A., von Cramon, D.Y., & Schubotz, R.I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience, 4*, 1–15. doi:10.3389/fnhum.2010.00025

Buhusi, C.V., & Meck, W.H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience, 6*, 755-765.

Camazine, S., Deneubourg, J., Franks, N.R., Sneyd, J., Theraulaz, G., & Bonabeau, E.. (2001). *Self-Organization in Biological Systems*. Princeton, Oxford: Princeton University Press.

Capizzi, M., Sanabria, D., & Correa, A. (2012). Dissociating controlled from automatic processing in temporal preparation. *Cognition, 123*, 293–302.

Caramazza, A. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology, 14*(1), 177–208.

Casillas, M., & Frank, M. (2012). The development of predictive processes in children's discourse understanding. In Knauff, M., Pauen, M., Sebanz, N., & Wachsmuth, I. (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 299–304). Austin, TX: Cognitive Society.

Caspers, J.. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics, 31*, 251–276.

Chater, N., & Manning, C.D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences, 10*(7), 335–344. doi:10.1016/j.tics.2006.05.006

Christiansen, M.H., & Chater, N. (2001). Connectionist psycholinguistics: capturing the empirical data. *Trends in Cognitive Sciences, 5*(2), 82–88. doi:10.1016/S1364-6613(00)01600-4

Clark, H.H. (1979). Responding to indirect speech acts. *Cognitive Psychology, 11*(4), 430–477.

Clark, H.H. (1996). *Using language.* Cambridge: Cambridge University Press.

Clark, H.H., & Wilkes-Gibbs, D. (1986). Reffering as a collaborative process. *Cognition, 22*, 1–39.

Clark, H.H., & Fox Tree, J.E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition, 84*, 73-111.

Coull, J.T., & Nobre, A.C. (1998). Where and When to Pay Attention: The Neural Systems for Directing Attention to Spatial Locations and to Time Intervals as Revealed by Both PET and fMRI. *The Journal of Neuroscience, 18*(18), 7426–7435.

Couper-Kuhlen, E. (2012). Some truths and untruths about final intonation in conversational questions. In J. P. De Ruiter (Ed.), *Formal, functional and interactional perspectives* (pp. 123–145). Cambridge, UK.: Cambridge University Press.

Crawley, M.J. (2009a). *The R Book*. England: John Wiley & Sons Ltd.

Cutler, A., & Pearson, M. (1986). On the analysis of prosodic turn-taking cues. In *Intonation and discourse* (pp. 139–155). London: Croom Helm.

Damian, M.F., Vigliocco, G., & Levelt, W.J.M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition, 81*, B77–B86.

Davidson, J.. (1984). Subsequent versions of invitations, offers, requests, and proposals dealing with potential or actual rejection. In J.M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 102–128.). Cambridge: Cambridge University Press.

De Ruiter, J.P. (2013). Methodological paradigms in interaction research. In I. Wachsmuth, J.P. De Ruiter, P. Jaecks, & S. Kopp (Eds.), *Alignment in Communication: Towards a New Theory of Communication* (pp. 11–31). Amsterdam: John Benjamins.

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

Dell, G.S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review, 93*(3), 283–321.

DeLong, K.A., Urbach, T.P., Groppe, D.M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology, 48*(9), 1203–1207.

DeLong, K.A., Urbach, T.P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121.

Duncan, S. (1974). On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, *3*, 161–180.

Duncan, S., & Fiske, D.W. (1977). *Face-to-face Interaction: Research, methods and theory.* Hillsdale, NJ: Lawrence Erlbaum.

Egorova, N., Pulvermüller, F., & Shtyrov, Y.. (2013). Neural Dynamics of Speech Act Comprehension: An MEG Study of Naming and Requesting. *Brain Topography.* doi:10.1007/s10548-013-0329-3

Esser, J., & Polomski, A. (1988). *Comparing reading and speaking intonation.* Amsterdam: Rodopi B.V. Editions.

Ford, C.E., & Thompson, S.E. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management for turns. In Ochs, Elinor & Schegloff, Emanuel A. (Eds.), *Interaction and grammar.* Cambridge: Cambrdige University Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. doi:10.1038/nrn2787

Fuster, J.M. (2001). The Prefrontal Cortex - An Update: Time is of the Essence. *Neuron*, *30*, 319–333.

Gambi, C., & Pickering, M.J. (2011). A cognitive architecture for the coordination of utterances. *Frontiers in Psychology.* doi:doi: 10.3389/fpsyg.2011.00275

Garrod, S., & Pickering, M.J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.

Gehrig, J., Wibral, M., Arnold, C., & Kell, C.A. (2012). Setting up the speech production network: oscillaitons contribute to lateralized information routing. *Frontiers in Psychology.*

Geluykens, R.. (1988). On the myth of rising intonation in polar questions. *Journal of Pragmatics*, *12*(4), 467–485.

Gergely, Gy. (2001). The obscure object of desire: "nearly, but clearly not, like me": contingency preference in normal children versus children with autism. *Bulletin of the Menninger Clinic, 65*(3), 411–426.

Gísladóttir, R.S., Chwilla, D., Schriefers, H. & Levinson, S.C. (2012). Speech act recognition in conversation: Experimental evidence. Poster presented at *34th Annual Meeting of Cognitive Science Society*, Sapporo, Japan.

Gleitman, L.R., January, D., Nappa, R. & Trueswell, J.C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory & Language, 57*(4), 544-569.

Grandke, T. (1983). Interpolation Algorithms for Discrete Fourier Transforms of Weighted Signals. *IEEE Transactions on Instrumentation and Measurement, 32*(2), 350–355.

Griffin, Z.M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274-279.

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics, 21*, 501-529.

Grosjean, F., & Hirt, C. (1996). Using Prosody to Predict the End of Sentences in English and French: Normal and Brain-damaged Subjects. *Language and Cognitive Processes, 11*(1), 107–134.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences, 98*(2), 694–699. doi:10.1073/pnas.98.2.694

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences, 9*(9), 416–423. doi:10.1016/j.tics.2005.07.004

Hagoort, P., & Poeppel, D. (2013). The Infrastructure of the Language-Ready Brain. In M.A. Arbib (Ed.), *Language, Music and the Brain: A mysterious relationship* (pp. 233-255). Cambridge, MA: MIT Press.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology, 28*, 136-141.

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience, 37,* 347-62.

Healey, P.G.T, Purver, M., & Howes, C. (2014). Divergence in Dialogue. *PLoS ONE, 9*(6), e98598.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*, 555–568.

Heritage, J., Robinson, J., Eliott, M., Beckett, M., & Wilkes, M. (2007). Reducing patients' unmet concerns in primary care: the difference one word can make. *Journal of General Internal Medicine*, *22*(10), 1429–1433.

Horton, W.S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91-117.

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

Indefrey. P., & Levelt, W.J.M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.

Jaeger, T.F.. (2008). Categorical data analysis: Away from ANOVA's (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Jasper, H., & Penfield, W. (1949). Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus. *Archiv Für Psychiatrie Und Zeitschrift Neurologie*, *183*, 163–174.

Jefferson, G. (1986). Notes on "latency" in overlap onset. *Human Studies*, *9*, 153–183.

Jensen, O., & Mazaheri, A. (2010). Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Frontiers in Human Neuroscience*, *4*(186). doi:10.3389/fnhum.2010.00186

Jescheniak, J.D., & Levelt, W.J.M. (1994). Word frequency effects in speech production: Retreival of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843.

Jescheniak, J.D., Schriefers, H., & Hantsch, A. (2003). Utterance Format Affects Phonological Priming in the Picture-Word Task: Implications for Models of

Phonological encoding in Speech Production. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 441–454.

Kamide, Y., Altmann, G.T.M., & Haywood, S.L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica, 26*, 22–63.

Kendrick, K.H., & Torreira, F. (2014). The timing and construction of preference: A quantitative study. *Discourse Processes*. Advance Online Publication. doi:10.1080/0163853X.2014.955997.

Knoeferle, P., Croecker, M.V., Scheepers, C., & Pickering, M.J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition, 95*, 95–127.

Koester, D., & Schiller, N.O. (2008). Morphological priming in overt language production: Electrophysiological evidence from Dutch. *Neuroimage, 42*.

Kolk, H.H.J., Chwilla, D.J., van Herten, M., & Oor, P.J.W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*(1), 1–36.

Konopka, A. E. & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology, 73*, 1-40.

Kuhlen, A.K., & Brennan, S.E. (2013). Language in dialogue: when confederates might be hazardous to your data. *Psychonomic Bulletin and Review, 20*(1), 54–72.

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

Ladefoged, P., Silverstein, R., & Papcun, G. (1973). Interruptibility of speech. *The Journal of the Acoustical Society of America, 54*(4), 1105–1108.

Lancaster, J.L., Rainey, L.H., Summerlin, J.L., Freitas, C.S., Fox, P.T., Evans, A.C., … Mazziotta, J.C. (1997). Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping, 5*(4), 238–242. doi:10.1002/(SICI)1097-0193(1997)5:4<238::AID-HBM6>3.0.CO;2-4

Lau, E.F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920–933. doi:10.1038/nrn2532

Levelt, W.J.M., & Kelter, S. (1982). Surface Form and Memory in Question Answering. *Cognitive Psychology*, *14*, 78–106.

Levelt, Willem J.M. (1989). *Speaking: From intention to articulation.* Cambridge, MA.: MIT Press.

Levelt, Willem J.M., Schriefers, H., Vorberg, D., Meyer, A.S., Pechmann, T., & Havinga, J. (1991). The time course of lexcial access in speech production: A study of picture naming. *Psychological Review, 98*(1), 122–142.

Levelt, W.J.M. & Meyer, A. S. (2000). Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology, 12*(4), 433-452.

Levin, H., Schaffer, C.A., & Snow, C. (1982). The prosodic and paralinguistic features of reading and telling stories. *Language and Speech, 25.*

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

Levinson, S.C. (2013). Action formation and ascription. In Stivers, Tanya & Sidnell, Jack (Eds.), *The handbook of conversation analysis* (pp. 103–130). Malden, MA.: Wiley-Blackwell.

Lewis P.A., & Miall, R.C. (2003). Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia, 41*(12), 1583–1592.

Local, J.K., Wells, B., & Seba, M. (1985). Phonetic aspects of turn delimination in London Jamaican. *Journal of Pragmatics, 9*, 309–330.

Local, J.K., Kelly, J., & Wells, B. (1986). Towards a phonology of conversation: Turn-taking in Tyneside English. *Journal of Linguistics, 22*, 411–437.

Local, J.K., & Walker, G. (2012). How phonetic features project more talk. *Journal of the International Phonetic Association, 42*(3), 255–280.

Logan, G.D., & Cowan, W.B. (1984). On the Ability to Inhibit Thought and Action: A Theory of an Act of Control. *Psychological Review*, *91*(3), 295–327.

Macar, F., Lejeune, H., Bonnet, M., Ferrara, A., Pouthas, V., Vidal, F., & Maquet, P. (2002). Activation of the supplementary motor area and of attentional networks during temporal processing. *Experimental Brain Research*, *142*, 475–485.

MacLeod, C.M., & MacDonald, P.A. (2000). Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*(10), 383–391. doi:10.1016/S1364-6613(00)01530-8

Magyari, L., & De Ruiter, J.P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, *3*. doi:doi: 10.3389/fpsyg.2012.00376

Manning, C.D., & Schütze, C.T. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA.: MIT Press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024

Max Planck Institute for Psycholinguistics. (2001). WebCelex. *WebCelex*. Retrieved from http://celex.mpi.nl/

Mehl, M.R., Vazire, S., Ramirez-Esparza, N., Slachter, R.B., & Pennebaker, J.W. (2007). Are Women Really More Talkative Than Men? *Science*, *317*(82).

Meyer, A.S. (1996). Lexical Access in Phrase and Sentence Production: Results from Picture-Word Interferece Experiments. *Journal of Memory and Language*, *35*, 477–496.

Mitra, P.P., & Pesaran, B. (1999). Analysis of Dynamic Brain Imaging Data. *Biophysical Journal*, *76*(2), 691–708. doi:10.1016/S0006-3495(99)77236-X

Müller-Gethmann, H., Ulrich, R., & Rinkenauer, G. (2003). Locus of the effect of temporal preparation: Evidence from the lateralized readiness potential. *Psychophysiology*, *40*.

Niemi, P., & Näätänen, R. (1981a). Foreperiod and Simple Reaction Time. *Psychological Bullettin*, *89*, 133–162.

Nobre, A.C., Correa, A., & Coull, J.T. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*, 1–6.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience, 2011*, 1–9. doi:10.1155/2011/156869

Oostenveld, R., Praamstra, P., Stegeman, D., & van Oosterom, A. (2001). Overlap of attention and movement-related activity in lateralized event-related brain potentials. *Clinical Neurophysiology*, *112*(3), 477–484. doi:10.1016/S1388-2457(01)00460-6

Papoutsi, M., de Zwart, J.A., Jansma, J.M., Pickering, M.J., Bednar, J.A., & Horwitz, B. (2009). From Phonemes to Articulatory Codes: An fMRI Study of the Role of Broca's Area in Speech Production. *Cerebral Cortex*, *19*, 2156–2165.

Pfurtscheller, G., & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and Clinical Neurophysiology*, *42*(6), 817–826.

Pfurtscheller, G., & Lopes da Silva, F.H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, *110*, 1842–1857.

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.M., & Bonnefond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PLoS ONE*, *9*(2). doi:10.1371/journal.pone.0088674

Piai, V., Roelofs, A., & Van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. *Brain Research*, *1450*, 87–101.

Pickering, M.J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169-226.

Pickering, M.J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110.

Pickering, M.J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral Brain Sciences*, *36*(4), 329–347.

Pomerantz, A. (1984). Agreeing and disagreeing with assesments: some features of preferred/dispreffered turn shapes. In J.M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 53–101.). Cambridge: Cambridge University Press.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582.

R Development Core Team. (2009). R: A language and environment for statistical computing. Retrieved from http://www.R-project.org

Rektor, I., Sochůrková, D., & Bočková, M. (2006). Intracerebral ERD/ERS in voluntary movement and in cognitive visuomotor task. In *Event-related dynamics of brain oscillations* (Vol. 159, pp. 311–330). Amsterdam: Elsevier.

Rohenkohl, G., Coull, J.T., & Nobre, A.C. (2011). Behavioural Dissociation between Exogenous and Endogenous Temporal Orienting of Attention. *PLoS ONE*, *6*(1). doi:doi:10.1371/journal.pone.0014620

Rommers, J., Dijkstra, T., & Bastiaansen, M.C.M. (2013). Context-dependent Semantic Processing in the Human Brain: Evidence from Idiom Comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762–776.

Rommers, J., Meyer, A.S., Piai, V., & Huettig, F. (2013). Constraining the involvement of language production in comprehension: A comparison of object naming and object viewing in sentence context. In *Seeing what's next: Processing and anticipating language referrring to objects*. Nijmegen: Max Planck Institute for Psycholinguistics.

Rossano, F. (2012). *Gaze behavior in face-to-face interaction*. Radboud Universiteit Nijmegen, Nijmegen.

Rossano, F., Brown, P., & Levinson, S.C. (2009). Gaze, questioning and culture. In *Conversation analysis: Comparative perspectives* (pp. 187–249.). Cambridge: Cambridge University Press.

Saarinen, T., Laaksonen, H., Parviainen, T., & Salmelin, R. (2006). Motor Cortex Dynamics in Visuomotor Production of Speech and Non-speech Mouth Movements. *Celebral Cortex*, *16*, 212–222.

Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

Salmelin, R., Schnitzler, A., Schmitz, F., & Freund, H.J. (2000). Single word reading in developmental stutterers and fluent speakers. *Brain, 123*, 1184–202.

Schaffer, D. (1983). The role of intonation as a cue to turn taking in conversation. *Journal of Phonetics, 11*, 243–257.

Schegloff, E.A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society, 29*, 1–63.

Schegloff, E.A. (1996). Turn organization: one intersection of grammar and interaction. In E. Ochs, E.A. Schegloff, & S.A. Thompson (Eds.), *Interaction and grammar* (pp. 52–133.). Cambridge: Cambridge University Press.

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

Schnurr, T.T., Costa, A., & Caramazza, A. (2006). Planning at the Phonological Level during Sentence Production. *Journal of Psycholinguistics Research, 35*(2), 189–213.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and mind moving together. *Trends in Cognitive Sciences, 10*(2), 70–76.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 76*, 379–423. doi:10.1145/584091.584093

Sosnik, R., Shemesh, M., & Abeles, M. (2007). The point of no return in planar hand movements: an indication of the existence of high level motion primitives. *Cognitive Neurodynamics, 1*, 341–358.

Stivers, T. (2010). An overview of the question-response system in American English conversation. *Journal of Pragmatics, 42*, 2772–2781.

Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S.C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106.*(26.), 10587–10592.

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking Lexical Access in Speech Production: Electrophysiological Correlates of Word Frequency and Cognate Effects. *Cerebral Cortex, 20*, 912–928.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268(5217),* 1632.

Ten Bosch, L., Oostdijk, L., & De Ruiter, J.P. (2005). Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues. *Presented at the 7th International Conference of Neural Information Processing ICONIP.*

Tremblay, P., Schiller, D.M., & Gracco, V.L. (2008). On the time-course and frequency selectivity of the EEG for different modes of response selection: Evidence form speech production and keyboard pressing. *Clinical Neurophysiology*, *119*, 88–99.

Trillenberg, P., Verleger, R., Wascher, E., Wauschkuhn, B., & Wessel, K. (2000). CNV and temporal uncertainty with "ageing" and "non-ageing" S1-S2 intervals. *Clinical Neurophysiology*, *111*(7), 1216–1226.

Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467.

Van Berkum, J.J.A. (2013). Anticipating communication. *Theoretical Linguistics, 39(1-2)*, 75-86.

Van de Meerendonk, N., Indefrey. P., Chwilla, D.J., & Kolk, H.H.J. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *Neuroimage*, *54*, 2350–2363.

Van Turennout, M., Hagoort, P., & Brown, C.M. (1998). Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science*, *280*, 572–574.

Von Hofsten, C., Uhlig, H., Adell, M., & Kochukhova, O. (2009). How children with autism look at events. *Research in Autism Spectrum Disorders*, *3*, 556–569.

Wagner, V., Jescheniak, J.D., & Schriefers, H. (2010). Phonological advance planning during sentence production: Effects of cognitive load on multiple access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36*, 323-340.

Wang, L., Zhu, Z., & Bastiaansen, M.C.M. (2012). Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. *Frontiers in Psychology*, *3*(187). doi:doi: 10.3389/fpsyg.2012.00187

Weiss, S., & Mueller, H.M. (2012). "Too many betas do not spoil the broth": the role of beta brain oscillations in language processing. *Frontiers in Psychology, 3*(201). doi:doi: 10.3389/fpsyg.2012.00201

Weiss, S., Mueller, H.M., Schack, B., King, J.W., Kutas, M., & Rappelsberger, P. (2005). Increased neuronal communication accompanying sentence comprehension. *International Journal of Psychophysiology, 57*, 129–141.

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication, 63-64*, 38-54.

Wicha, N.Y.Y., Moreno, E.M., & Kutas, M. (2004). Anticipating words and their gender: an event-related brain potentials study of semantic integration, gender expectancy and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288.

Willems, R.M., De Boer, M., De Ruiter, J.P., Noordzij, M.L., Hagoort, P., & Toni, I. (2010). A Dissociation Between Linguistic and Communicative Abilities in the Human Brain. *Psychological Science, 21*(1), 8–14.

Wilson, M., & Wilson, T.P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review, 12*(6), 957–968.

8 Samenvatting

Dit proefschrift onderzoekt de cognitieve processen die plaatsvinden tijdens dagelijkse, natuurlijke gesprekken. Meer specifiek presenteert het een mogelijke verklaring voor hoe sprekers hun beurten timen. Dagelijkse gesprekken verlopen normaliter gladjes en moeiteloos. Een nadere beschouwing van het tijdsverloop van gesprekken veronderstelt echter dat er een complex cognitief netwerk schuilgaat onder vloeiend verbaal taalgebruik. De volgende hypothese vormt het startpunt van dit proefschrift: Sprekers voorspellen wanneer een beurt eindigt door te voorspellen hoe deze eindigt. Om dit te testen, worden in vier onderzoeken verschillende technieken gebruikt, van het drukken op een knop en het meten wanneer iemand begint met praten tot gating en electrofysiologie.

De eerste twee onderzoeken bekijken of sprekers dichter bij het einde van de vorige beurt antwoorden (verbaal of door het drukken op een knop) als ze kunnen antciperen op het einde van die beurt. Het tweede onderzoek bekijkt ook of snelle beurt-overgangen alleen mogelijk zijn als de volgende spreker al vroeg tijdens de huidige beurt weet wat hij/zij gaat zeggen. De tweede twee hoofdstukken richten zich op de oscillerende hersencorrelaten van het voorspellen van beurt-eindes en van de voorbereiding van spraak.

Over het algemeen laten de gedragsresultaten van de onderzoeken zien dat wanneer sprekers proberen te antwoorden op het moment dat de beurt eindigt, ze beïnvloed worden door (1) de hoeveelheid voorspelde linguïstische informatie in de beurt en (2) hoe vroeg ze weten wat ze moeten antwoorden. Als het mogelijk is om de inhoud van de beurt met zekerheid te voorspellen, antwoorden sprekers sneller. Sprekers antwoorden ook sneller als ze eerder weten wat ze moeten antwoorden. Met betrekking tot de EEG resultaten, werd in hoofdstuk 4 een hersencorrelaat van het voorspellen van beurt-eindes, namelijk desynchronisatie in de beta frequentie, gelocaliseerd in frontale en linkse temporaal-parietale hersengebieden. In hoofdstuk 5 werd een strerkere alfa synchronisatie gevonden in linkse occipitale, parietale en temporale hersengebieden als het verbale antwoord voor een langere tijd werd uitgesteld. Dat zou kunnen betekenen dat alfa synchronizatie onderdrukking van de voorbereiding van spraak weerspiegelt op het moment dat sprekers vroeg weten wat ze moeten antwoorden maar pas later hun antwoord moeten geven. De gedrags- en EEG resultaten ondersteunen de hypothese dat voorspellingen van hoe een beurt eindigt de timing van beurtwisselingen in conversaties beïnvloeden.

9 Curriculum Vitae

Lilla Magyari studied Psychology and Hungarian Grammar and Literature at the ELTE University in Budapest, Hungary. She then moved to Nijmegen to study in the Cognitive Neuroscience Research Master's Programme. After obtaining her master's degree she started her Ph.D. in the Multimodal Interaction Project within the Language & Cognition Group of the Max Planck Institute for Psycholinguistics. Her dissertation explores the cognitive mechanisms involved in the timing of turn-taking in everyday conversations. Complementing her Ph.D. studies, she worked as an assistant in the Neurobiology of Language Group of the Max Planck Institute and in the Neuroimaging Center of the Donders Institute for Brain, Cognition and Behaviour, focusing on the implementation of methods for EEG/MEG data-analysis within the FieldTrip software-package. She also studied theatre-directing at the Amsterdam School of the Arts for a year. Currently, she lives in Budapest where she works as an assistant professor at the Department of General Psychology of the Pázmány Péter Catholic University. Her research investigates linguistic and cultural differences in turn-taking of natural conversation, empirical aesthetics and coordination of movement improvisation.

10 List of publications

Moreno, I., De Vega, M., León, I., Bastiaansen, M. C. M., Lewis, A. G., & Magyari, L. (2015). Brain dynamics in the comprehension of action-related language. A time-frequency analysis of mu rhythms. *Neuroimage,* 109, 50-62. doi:10.1016/j.neuroimage.2015.01.018.

Magyari, L., Bastiaansen, M. C. M., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience,* 26(11), 2530-2539. doi:10.1162/jocn_a_00673.

Magyari, L. (2014). Beszédprodukció [Speechproduction]. In Cs. Pléh & Á. Lukács (Eds.), *Pszicholingvisztika 1-2. Magyar pszicholingvisztikai kézikönyv* [Psycholinguistics 1-2. Hungarian handbook of psycholinguistics] (pp. 371-410). Budapest: Akadémiai Kiadó.

Drew, P., Floyd, S., Gipper, S., Gisladottir, R. S., Hoymann, G., Kendrick, K. H., Levinson, S. C., Magyari, L., Manrique, E., Rossi, G., San Roque, L., & Torreira, F. (2013). Huh? What? – A first survey in 21 languages. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational repair and human understanding* (pp. 343-380). New York: Cambridge University Press.

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

Wang, L., Jensen, O., Van den Brink, D., Weder, N., Schoffelen, J.-M., Magyari, L., Hagoort, P., & Bastiaansen, M. C. M. (2012). Beta oscillations relate to the N400m during language comprehension. *Human Brain Mapping,* 33, 2898-2912. doi:10.1002/hbm.21410

Bastiaansen, M. C. M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience,* 22, 1333-1347. doi:10.1162/jocn.2009.21283

Magyari, L. (2008). A mentális lexikon modelljei és a magyar nyelv [Models of the mental lexicon and the Hungarian language]. In J. Gervain, & C. Pléh (Eds.), *A láthatatlan nyelv* [Invisible Language]. Budapest: Gondolat Kiadó.

Magyari, L., & De Ruiter, J. P. (2008). Timing in conversation: The anticipation of turn endings. In J. Ginzburg, P. Healey, & Y. Sato (Eds.), *Proceedings of the 12th Workshop on the Semantics and Pragmatics Dialogue* (pp. 139-146). London: King's college.

Magyari, L. (2005). A nyelv miért nem olyan, mint a szem? [Why is language not like vertebrate eye?]. In J. Gervain, K. Kovács, Á. Lukács, & M. Racsmány (Eds.), *Az ezer arcú elme* [The mind with thousand faces] (first edition, pp. 452-460). Budapest: Akadémiai Kiadó.

11 MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. Miranda van Turennout
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. Niels O. Schiller
3. Lexical access in the production of ellipsis and pronouns. Bernadette M. Schmitt
4. The open-/closed-class distinction in spoken-word recognition. Alette Haveman
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. Kay Behnke
6. Gesture and speech production. Jan-Peter de Ruiter
7. Comparative intonational phonology: English and German. Esther Grabe
8. Finiteness in adult and child German. Ingeborg Lasser
9. Language input for word discovery. Joost van de Weijer
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. James Essegbey
11. Producing past and plural inflections. Dirk Janssen
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. Anna Margetts
13. From speech to words. Arie van der Lugt
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. Eva Schultze-Berndt
15. Interpreting indefinites: An experimental study of children's language comprehension. Irene Krämer
16. Language-specific listening: The case of phonetic sequences. Andrea Weber
17. Moving eyes and naming objects. Femke van der Meulen
18. Analogy in morphology: The selection of linking elements in Dutch compounds. Andrea Krott
19. Morphology in speech comprehension. Kerstin Mauth
20. Morphological families in the mental lexicon. Nivja H. de Jong
21. Fixed expressions and the production of idioms. Simone A. Sprenger
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). Birgit Hellwig
23. Paradigmatic structures in morphological processing: Computational and crosslinguistic experimental studies. Fermín Moscoso del Prado Martín
24. Contextual influences on spoken-word processing: An electro-physiological approach. Daniëlle van den Brink

25. Perceptual relevance of prevoicing in Dutch. Petra M. van Alphen
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. Joana Cholin
27. Producing complex spoken numerals for time and space. Marjolein Meeuwissen
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. Rachél J. J. K. Kemps
29. At the same time...: The expression of simultaneity in learner varieties. Barbara Schmiedtová
30. A grammar of Jalonke argument structure. Friederike Lüpke
31. Agrammatic comprehension: An electrophysiological approach. Marlies Wassenaar
32. The structure and use of shape-based noun classes in Mirãna (North West Amazon). Frank Seifart
33. Prosodically-conditioned detail in the recognition of spoken words. Anne Pier Salverda
34. Phonetic and lexical processing in a second language. Mirjam Broersma
35. Retrieving semantic and syntactic word properties. Oliver Müller
36. Lexically-guided perceptual learning in speech processing. Frank Eisner
37. Sensitivity to detailed acoustic information in word recognition. Keren B. Shatzman
38. The relationship between spoken word production and comprehension. Rebecca Özdemir
39. Disfluency: Interrupting speech and gesture. Mandana Seyfeddinipur
40. The acquisition of phonological structure: Distinguishing contrastive from noncontrastive variation. Christiane Dietrich
41. Cognitive cladistics and the relativity of spatial cognition. Daniel B.M. Haun
42. The acquisition of auditory categories. Martijn Goudbeek
43. Affix reduction in spoken Dutch. Mark Pluymaekers
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. Valesca Kooijman
45. Space and iconicity in German Sign Language (DGS). Pamela Perniss
46. On the production of morphologically complex words with special attention to effects of frequency. Heidrun Bien
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. Amanda Brown
48. The acquisition of verb compounding in Mandarin Chinese. Jidong Chen
49. Phoneme inventories and patterns of speech sound perception. Anita Wagner
50. Lexical processing of morphologically complex words: An information-theoretical perspective. Victor Kuperman

51. A grammar of Savosavo, a Papuan language of the Solomon Islands. Claudia Wegener
52. Prosodic structure in speech production and perception. Claudia Kuzla
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. Sarah Schimke
54. Studies on intonation and information structure in child and adult German. Laura de Ruiter
55. Processing the fine temporal structure of spoken words. Eva Reinisch
56. Semantics and (ir)regular inflection in morphological processing. Wieke Tabak
57. Processing strongly reduced forms in casual speech. Susanne Brouwer
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. Miriam Ellert
59. Lexical interactions in non-native speech comprehension: Evidence from electroencephalography, eye-tracking, and functional magnetic resonance imaging. Ian FitzPatrick
60. Processing casual speech in native and non-native language. Annelie Tuinman
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. Stuart Robinson
62. Evidentiality and intersubjectivity in Yurakar´e: An interactional account. Sonja Gipper
63. The influence of information structure on language comprehension: A neurocognitive perspective. Lin Wang
64. The meaning and use of ideophones in Siwu. Mark Dingemanse
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. Marco van de Ven
66. Speech reduction in spontaneous French and Spanish. Francisco Torreira
67. The relevance of early word recognition: Insights from the infant brain. Caroline Junge
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. Matthias J. Sjerps
69. Structuring language : contributions to the neurocognition of syntax. Katrien R. Segaert
70. Infants' appreciation of others' mental states in prelinguistic communication : a second person approach to mindreading. Birgit Knudsen
71. Gaze behavior in face-to-face interaction. Federico Rossano
72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. Connie de Vos
73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. Attila Andics

74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. Marijt Witteman
75. The use of deictic versus representational gestures in infancy. Daniel Puccini
76. Territories of knowledge in Japanese conversation. Kaoru Hayano
77. Family and neighbourhood relations in the mental lexicon: A cross-language per spective. Kimberley Mulder
78. Contributions of executive control to individual differences in word production. Zeshu Shao
79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual pro cessing. Patrick van der Zande
80. High pitches and thick voices: The role of language in space-pitch associations. Sarah Dolscheid
81. Seeing what's next: Processing and anticipating language referring to objects. Joost Rommers
82. Mental representation and processing of reduced words in casual speech. Iris Hanique
83. The many ways listeners adapt to reductions in casual speech. Katja Poellmann
84. Contrasting opposite polarity in Germanic and Romance languages: Verum focus and affirmative particles in native speakers and advanced L2 learners. Giuseppina Turco
85. Morphological processing in younger and older people: Evidence for flexible dualroute access. Jana Reifegerste
86. Semantic and syntactic constraints on the production of subject-verb agreement. Alma Veenstra
87. The acquisition of morphophonological alternations across languages. Helen Buckler
88. The evolutionary dynamics of motion event encoding. Annemarie Verkerk
89. Rediscovering a forgotten language. Jiyoun Choi
90. The road to native listening: Language-general perception, language-specific input. Sho Tsuji
91. Infants' understanding of communication as participants and observers. Gudmundur Bjarki Thorgrímsson
92. Information structure in Avatime. Saskia van Putten
93. Switch reference in Whitesands. Jeremy Hammond
94. Machine learning for gesture recognition from videos. Binyam Gebrekidan Gebre
95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. Beyza Sumer

96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. Salomi Savvatia Asaridou
97. Incrementality and Flexibility in Sentence Production. Maartje van de Velde
98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. Edwin van Leeuwen
99. The request system in Italian interaction. Giovanni Rossi
100. Timing turns in conversation: A temporal preparation account. Lilla Magyari