



1- and 2-year-olds' expectations about third-party communicative actions



Gudmundur B. Thorgrimsson^{a,*}, Christine Fawcett^b, Ulf Liszkowski^{a,c}

^a Max-Planck Research Group, Communication Before Language, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 Nijmegen, Netherlands

^b Department of Psychology, Uppsala University, Von Kraemers allé 1, S-751 42 Uppsala, Sweden

^c Department of Developmental Psychology, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany

ARTICLE INFO

Article history:

Received 13 August 2014

Received in revised form 19 January 2015

Accepted 10 February 2015

Keywords:

Action understanding

Third-party interactions

Turn-taking

Communicative development

Eye tracking

ABSTRACT

Infants expect people to direct actions toward objects, and they respond to actions directed to themselves, but do they have expectations about actions directed to third parties? In two experiments, we used eye tracking to investigate 1- and 2-year-olds' expectations about communicative actions addressed to a third party. Experiment 1 presented infants with videos where an adult (the Emitter) either uttered a sentence or produced non-speech sounds. The Emitter was either face-to-face with another adult (the Recipient) or the two were back-to-back. The Recipient did not respond to any of the sounds. We found that 2-, but not 1-year-olds looked quicker and longer at the Recipient following speech than non-speech, suggesting that they expected her to respond to speech. These effects were specific to the face-to-face context. Experiment 2 presented 1-year-olds with similar face-to-face exchanges but modified to engage infants and minimize task demands. The infants looked quicker to the Recipient following speech than non-speech, suggesting that they expected a response to speech. The study suggests that by 1 year of age infants expect communicative actions to be directed at a third-party listener.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

The ability to understand and anticipate the actions of other people is an essential part of human development. The rudiments of this ability are found in infancy, enabling infants to learn from and communicate with others before they develop the means to participate in linguistic exchanges. Infants monitor others' actions with interest and understand much about what goes on when people act on their environment. They understand that actions can be directed at objects (Behne, Carpenter, & Tomasello, 2005; Woodward, 1998), and they infer unseen goals of ongoing actions (Nielsen, 2009; Southgate & Csibra, 2009; Southgate, Johnson, El Karoui, & Csibra, 2010). Recent eye tracking studies show that infants also make visual anticipations to the goal location of actions, such as an arm transferring a ball into a bucket (Falck-Ytter, Gredebäck, & von Hofsten, 2006), an actor bringing a cup to her mouth (Hunnus & Bekkering, 2010), or a spoon to another person's mouth (Gredebäck & Melinder, 2010).

* Corresponding author. Tel.: +31 24 3521490; fax: +31 24 3521213.
E-mail address: gudmundurbjarki@gmail.com (G.B. Thorgrimsson).

Of course, not all human actions involve an observable physical movement toward a goal object. People act on each other remotely through speech, gesture, and other forms of communicative actions. Infants are surrounded with the communicative interactions of others, and observing and overhearing third-party interactions forms a large part of infants' communicative experience, especially in communities where it is less common for caregivers to directly address preverbal infants (Brown, 1998; Heath, 1983; Ochs & Schieffelin, 1984, see also Lieven, 1994). How much do infants understand about others' interactions? Do they understand communicative actions as attempts to influence other people (i.e. communication as person-directed action)? Several studies indicate that infants in their second year can learn through observing object-directed actions that communicate information to a third party. For example, at 14 months, infants can infer the location of a hidden object from a pointing gesture addressed to a third party (Gräfenhain, Behne, Carpenter, & Tomasello, 2009). At 18 months they can learn novel object labels from third-party labeling (Floor & Akhtar, 2006), and learn from object-directed emotions displayed to a third party (Repacholi & Meltzoff, 2007). These studies reveal that infants understand and make use of information conveyed through communicative actions that are not directed to them, demonstrating that infants' understanding extends beyond their own interactions. It is less clear, however, what infants understand of the structure of the interaction itself, that is, whether they expect communicative actions to be directed at interactants. Attending to and anticipating how people respond to communication in third-party interactions would facilitate social-observational learning, and provide evidence that infants' understanding of communicative exchanges extends beyond the second-person understanding manifested in their own interactions (Reddy, 2008).

Recent evidence suggests that infants monitor and understand third-party gestural interactions and can anticipate their outcome. 14-month-old infants viewing a person point or direct a request gesture (open hand, palm up) to an object, successfully anticipated that the addressee would transfer the object to the gesturer (Thorgrimsson, Fawcett, & Liskowski, 2014). Infants also appear to have some expectations toward third-party addressees of speech. Looking time measures suggest that infants expect a person addressed with a nonsense word to select the same object the speaker had previously expressed a preference for (unbeknownst to the addressee; Martin, Onishi, & Vouloumanos, 2012). These findings suggest that infants may understand speech as an action directed to another person to provoke a response, enabling them to actively anticipate how third-party verbal interactions unfold.

A recent line of eye-tracking studies has examined infants' looking patterns while viewing third-party conversations. Infants begin to visually orient toward third-party speakers in the second half of their first year and orienting to speakers or other sources of sound becomes more flexible over the first three years. Thus, 6- and 11-month-olds viewing a third-party conversation shift their gaze from the previous to the current speaker during roughly half of the conversational turns, exhibiting rudimentary abilities to follow a third-party turn-taking conversation (Augusti, Melinder, & Gredebäck, 2010). By 3 years of age, children also disengage from the current speaker and orient their gaze to the prospective speaker during the majority of turns (von Hofsten, Uhlig, Adell, & Kochukhova, 2009). In the latter study, 3-year-olds, but not 1-year-olds, were also found to make more such gaze shifts between human interlocutors than between turn-taking objects presented in a matched control condition. A recent follow-up study (Bakker, Kochukhova, & von Hofsten, 2010) similarly found that for 3-year-olds, but not 1-year-olds, such gaze-shifts were more frequent between people than between objects. Further, the older age group made more shifts between people or objects when these emitted speech than when they emitted mechanical sounds. Together, these eye tracking studies suggest that before 12 months of age, infants tend to fixate speakers or other sound sources during their turn, and that by 3 years, children are more flexible in monitoring others' interactions in that they disengage attention from human speakers to addressees.

However, the extent to which infants visually monitor speakers in a conversation does not necessarily reflect their expectations about responses to speech. Since these studies all presented contingent, alternating conversational turns, the infants may have shifted their gaze from speaker to listener (or between turn-taking objects) because they learned to anticipate the contingency of the stimuli, but not because they expected the utterances to provoke a response from the listener. This possibility is especially evident considering that from as early as 2 months of age, infants quickly learn to anticipate side-alternating visual events (Canfield & Haith, 1991). Thus, the fact that 1-year-olds shift their gaze to a contingently reacting object as much as to a human listener in a conversation (von Hofsten et al., 2009), may reflect only their sensitivity to the matched contingencies. Demonstrating that infants switch their gaze to a third-party addressee in the absence of previous or ongoing contingent interaction would give stronger evidence that they expect speech to provoke a response from another person. In addition, since infants are more likely to identify a person as the source of familiar natural human sounds than unfamiliar artificial sounds, assessing infants' expectations about speech will benefit from a comparison with their expectations about common non-communicative human sounds, such as coughs or yawns.

1.1. *The current study*

The current study built on previous research to examine infants' expectations about communicative actions in more detail. We investigated whether infants expect an utterance directed to a third party to be responded to in the absence of previous contingent interactions. To contrast communicative with non-communicative actions, infants were also presented with identical videos in which the Emitter made a natural non-speech sound (e.g. a cough). We tested 12- and 24-month-olds to investigate whether both older, verbal infants and younger infants with much less linguistic experience would have expectations about speech. Infants at the age of 12 months are only beginning to utter their first words,

but during the first year they typically have much experience not only observing and overhearing others' communication, but also participating in so-called protoconversations – face-to-face interactions where they engage directly with others' speech (Bateson, 1975). These experiences might contribute to the development of infants' expectations about third-party communication.

In the study, we presented infants with videos of an adult actor in profile (the *Emitter*) uttering a short sentence in the presence of another adult actor (the *Recipient*). Crucially, the Recipient never produced a response, ruling out the possibility that infants could anticipate responses based only on their contingency. We reasoned that if infants expect third parties to respond to speech they should do so in the current situation even without having seen a response occur, and possibly from the first trial on. Previous studies suggest that infants as young as 4 months of age have expectations about event outcomes despite having had no previous experimental exposure to the event (Wang, Baillargeon, & Brueckner, 2004). Moreover, eye-tracking studies have shown that despite repeated viewings, infants do not learn to anticipate unusual outcomes of familiar actions (such as a person bringing a spoon to her ear: Hunnius & Bekkering, 2010), suggesting that their action expectations are established outside of the experimental setting.

We also manipulated the orientation of the actors. As mutual gaze between third parties signals affiliation for preschoolers (Nurmsoo, Einav, & Hood, 2012) and adults (Thayer & Schiff, 1974), and speakers typically face their addressees, we expected the body orientation and gaze-direction of the actors to influence infants' expectations. Further, 6-month-olds monitor interlocutors more when they are in mutual than non-mutual gaze (Augusti et al., 2010), and 10-month-olds not only distinguish mutual gaze from non-mutual gaze but also show signs of surprise when an entity that a person looks at and converses with is revealed to be an object, rather than another person (Beier & Spelke, 2012). Studies on infants' understanding of object-directed actions further suggest that gaze direction is a strong cue to goal expectation (Phillips, Wellman, & Spelke, 2002). Thus, it is possible that infants understand speech to be communicative or provoke a response only when the speaker is facing her addressee. In Experiment 1, the actors either were in a typical conversational context with face-to-face body orientation and mutual gaze, or were facing away from each other in a back-to-back configuration (as in Augusti et al., 2010). In Experiment 2 we maintained a face-to-face body orientation but one in which the actors averted gaze and looked down at a paper. In addition, with a between-subject design, Experiment 2 allowed us to analyze how the lack of a contingent response influenced infants' expectations across repeated trials in the same condition. Together, the current study thus allowed an assessment of whether 12- and 24-month-old infants have expectations about third-party communicative actions, and whether these expectations are specific to speech and modulated by body orientation, in the absence of contingent responses.

We were interested in three main comparisons. Our main question was whether infants have stronger expectations of a response to speech than non-speech sounds when interactants are in a typical face-to-face configuration. In addition, we were interested in whether infants' expectations in the speech condition would remain in situations in which the interactants did not look at each other. Finally, we wanted to know whether infants have developed expectations about speech only by the time they are already verbal themselves at 24 months of age, or whether these expectations are already in place at 12-months, when infants have had very limited verbal experience. Our two dependent measures were infants' latency to orient their gaze from the Emitter to the Recipient, and the total time that infants looked at the Recipient after the sounds, with the general hypothesis that earlier and longer looking to the Recipient indicates a stronger expectation for her to respond. We tested two specific hypotheses. Our first hypothesis was that when the two actors faced each other, infants would be more likely to expect a response following speech than non-speech, as this is a typical communicative context. Second, given infants' early sensitivity to third-party mutual gaze (Augusti et al., 2010; Beier & Spelke, 2012), and their ability to use gaze to predict actions (Phillips et al., 2002), we predicted that infants would expect a response from speech only when the Emitter is oriented toward the Recipient. Regarding age, one possibility is that infants' understanding of third-party verbal interactions depends on their own verbal communicative experience (von Hofsten et al., 2009; see also Sommerville, Woodward, & Needham, 2005) and that 12-month-olds – who are only beginning to utter their first words – have not yet developed expectations about third-party speech. Alternatively, 12-month-olds' observations of others' conversations and their participation in proto-conversations may have already led them to form such expectations. To address this developmental question, we examined 12-month-olds' performance more closely in Experiment 2, using an optimized and simplified version of the paradigm used in Experiment 1.

2. Experiment 1

2.1. Method

2.1.1. Participants

Twenty-seven 12-month-olds (11 boys, 16 girls; mean age = 12;20, range = 11;24 to 12;29) and twenty-six 24-month-olds (17 boys, 9 girls; mean age = 24;18, range = 23;29 to 24;29) participated in the experiment. Counterbalancing was based on 24 infants per age group, but in case of dropouts or cancellations, additional infants had been scheduled to ensure the full sample size would be reached. These additional infants were retained to increase power. An additional 33 12-month-olds participated but were not included in the final sample: sixteen due to not meeting the looking criteria (i.e. looking at the Emitter during the period in which she emitted a sound in at least two trials per condition; see Section 2.1.5), nine due to not finishing the experiment as a result of fussiness, four due to technical error, and four due to unsuccessful calibration.



Fig. 1. A frame from the face-to-face condition (above) and the back-to-back condition videos (below) of Experiment 1. The rectangles define the two areas of interest.

For the 24-month-olds, an additional nine infants participated but were not included: seven due to not meeting the looking criteria, one due to not finishing the experiment as a result of fussiness, and one due to unsuccessful calibration. Infants were recruited from a database of families who expressed interest in participating in research. Infants were primarily white and from middle-class backgrounds, living in a medium-sized European city. Parents received a small gift for participating.

2.1.2. Apparatus

The infants' eye movements were measured using a Tobii 1750 remote eye-tracker with an infant add-on. The eye tracker is integrated with a 17" TFT monitor with a native resolution of 1280×1024 pixels. The Tobii 1750 has a 50 Hz sample rate, an accuracy of 0.5° , precision of 1° , and allows head movements of up to about 30 cm horizontally, 15 cm vertically, and 20 cm in depth.

2.1.3. Stimuli

Infants were presented with 24 videos (31.37×7.66 visual degrees), each with a duration of 6–7 s in total. The videos showed two actors in profile, visible from the shoulders up. Each video began with a period of inactivity (1 or 2 s), followed by the Emitter emitting a sound (mean duration 1.4 s), then by another period of inactivity (4 s). In half of the videos, the Emitter uttered speech and in the other half, she uttered a non-speech sound. The orientation of the two actors was also manipulated, such that in half of the videos, they were face-to-face and in mutual gaze and in the other half, they were back-to-back and gazing away from each other (see Fig. 1). The conditions were thus *face-to-face/speech*, *face-to-face/non-speech*, *back-to-back/speech*, and *back-to-back/non-speech*. The position and identity of the Emitter were counterbalanced across infants, such that both actors played the role of Emitter and Recipient but each actor was always on the same side of the screen. To avoid too much repetition across trials, three examples of each sound type were presented (see Table 1) and each of them was presented twice, with different durations of the initial inactive period (1 or 2 s). The aim was to make the trials less predictable and more likely to keep the infants engaged. Each condition was presented as a block of six trials with both block order and sound example order counterbalanced across infants.

Table 1

Characteristics of the speech and the non-speech sounds in Experiment 1.

Speech sound utterances	Non-speech sounds
"Hoi, hoe gaat het?" (Hi, how are you?)	Throat clearing
"Wil je wat eten?" (Do you want something to eat?)	Vocal sigh
"Ik ben moe vandaag" (I am tired today)	Song (syllable "ta" repeated five times at different pitch levels)
Mean duration = 1256.67 ms, <i>SD</i> = 174.32	Mean duration = 1580 ms, <i>SD</i> = 133.75

2.1.4. Procedure

Infants were seated in a safety car seat that was placed in their parents' lap so that the infants' eyes were approximately 60 cm from the monitor. Before the experiment, the infants' gaze was calibrated using a 9-point calibration during which the experimenter monitored the infants' attention on the screen and showed a short animated video of talking puppets in place of the calibration video to recapture attention when needed. Each infant viewed all four conditions of the experiment (3.5 min total). Between trials within a block, a short (1 s) animated video (a small abstract moving picture accompanied by sound) appeared in the middle of the screen in an attempt to ensure that the infants' gaze would be on the screen and centered at the onset of each trial; however, the onset of the trial was not contingent on infants' gaze to these short animations. Between trial blocks, similar but more engaging animated videos (a looming diamond shape accompanied by sound) were presented to center the infants' gaze and to help sustain their attention throughout the study. These longer videos were terminated by an experimenter when the infants' attention was on the screen or once the maximum duration of 9 s was reached. Audio was transmitted through a single desktop speaker connected to the computer and hidden from view behind the monitor.

2.1.5. Data reduction

Rectangular areas of interest (AOI) were created around each of the actors (7.98×8.65 visual degrees each; see Fig. 1). Infants' unfiltered gaze data points registered within these AOI's at a rate of 50 per second were used to calculate the dependent measures. To account for possible errors in gaze estimation, the AOI's covered an area approximately 30 pixels (0.8 visual degree) wider and higher than the actors (e.g. Gredebäck & Melinder, 2010). Two critical time phases were selected from the data: the sound phase, during which the Emitter made a sound, and the post-sound phase, during which both actors were silent and immobile. Since the time it takes for infants and adults to initiate a saccade is around 200 ms (Becker, 1972; Canfield, Smith, Brezsnjak, & Snow, 1997), each phase was offset by this amount of time. Thus, the sound phase began 200 ms after the onset of the sound and the post-sound phase began 200 ms after the end of the sound, resulting in a duration of 3.8 s for the latter. Two dependent measures – gaze shift latency and looking time – were extracted from the post-sound phase data. To ensure that the measures reflect looking behavior that is related to the sound event, trials in which no fixation was registered on the Emitter AOI during the sound phase were excluded from the analysis of both dependent measures. To measure these fixations to the Emitter during speech, we used the Tobii Clearview fixation filter, and classified gaze data points as fixations when they occurred within a radius of 30 pixels for at least 100 ms (e.g. Doi, Tagawa, & Shinohara, 2010; Hunnius, de Wit, Vriens, & von Hofsten, 2011). The fixation filter was only used for this one measure. All remaining data reduction is based on the unfiltered gaze data points. The number of excluded trials differed slightly between age groups and conditions: a 2 (sound type: speech vs. non-speech) \times 2 (orientation: face-to-face vs. back-to-back) \times 2 (age, between subjects) mixed analysis of variance (ANOVA) revealed that more trials were excluded in the speech conditions ($M = 1.915$, $SD = .997$) than in the non-speech conditions ($M = 1.475$, $SD = 1.020$), $F(1, 51) = 9.750$, $p = .003$, $\eta_p^2 = .160$. This difference is likely due to a slightly shorter duration of the speech sounds, which gave infants less time to orient to and fixate the Emitter. In addition, more trials were excluded per condition for the 12-month-olds ($M = 2.102$, $SD = .868$) than for the 24-month-olds ($M = 1.288$, $SD = .867$), $F(1, 51) = 11.648$, $p = .001$, $\eta_p^2 = .186$, possibly because the older age group was more attentive to the display overall. To reduce noise in the data, infants who fixated the Emitter in fewer than two trials of a condition were excluded from our analyses ($n = 16$). For within-subject designs, two trials per condition are considered the minimum to effectively reduce noise in similar experimental paradigms (see Fernald, Zangl, Portillo, & Marchman, 2008). Nevertheless, to ensure that these exclusions do not bias our results, we conducted an additional analysis on all infants who fixated the Emitter in at least one trial per condition.

The measure of gaze shift latency was used to capture how quickly the infants shifted their gaze from the Emitter to the Recipient following the Emitter's sound. Only trials that had a fixation on the Emitter during the sound could be considered. The measure was calculated by subtracting the start time of the post-sound phase from the time when the first gaze data point was registered within the Recipient AOI. In the main analyses, trials in which infants did not look at the Recipient during the post-sound phase were assigned a score of 3800 ms (the maximum possible latency). The benefits of assigning a maximum score are that these trials represent meaningful behavior (i.e. not looking to the AOI following the sound) and including them yields more data. However, to ensure that the results are not an artifact of these assignments, an additional analysis excluded trials without a look to the Recipient. Our second main measure was a looking time difference score to gauge how long the infants looked to the Recipient compared to the Emitter in the post-sound phase. Difference scores were used to help diminish individual variability in overall looking times. The difference score was calculated by subtracting the looking time to the Emitter from the looking time to the Recipient during the post-sound phase. Looking time was indexed by the number of gaze data points registered within each of the AOI's.

2.2. Results and discussion

2.2.1. Gaze shift latency

Our main question was whether infants looked quicker to and longer at the Recipient in a face-to-face context when she was addressed with speech compared to non-speech. In addition, we were interested in whether the orientation of the actors would have an effect on infants' looking following the sounds, and whether we would find evidence for specific expectations about speech already at 12 months of age. Fig. 2 displays the latency scores across conditions and age groups.

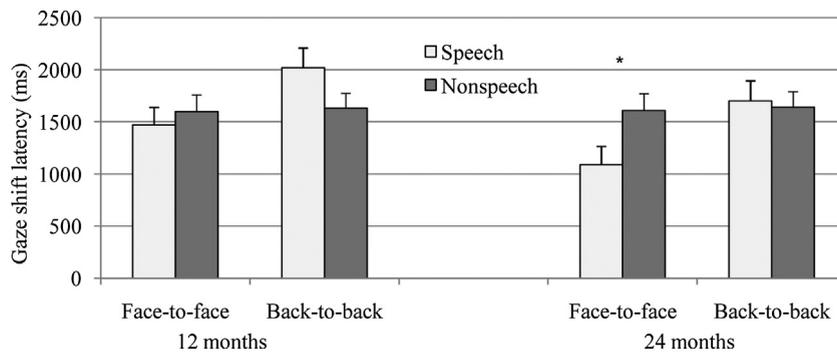


Fig. 2. Average gaze shift latency in Experiment 1 (untransformed), for all four conditions. Asterisks indicate a significant difference ($p < .05$).

The scores were square root transformed to a normal distribution for the analyses. A first omnibus test with a 2 (sound type: speech vs. non-speech) \times 2 (orientation: face-to-face vs. back-to-back) \times 2 (age: 12 vs. 24 months) mixed analysis of variance (ANOVA) revealed a main effect of orientation ($F(1,51) = 8.881$, $p = .004$, $\eta_p^2 = .148$), as well as interactions between orientation and sound ($F(1,51) = 6.276$, $p = .015$, $\eta_p^2 = .110$), and between sound and age ($F(1,51) = 4.868$, $p = .032$, $\eta_p^2 = .087$). To directly address our main questions, we compared latency for speech vs. non-speech in the face-to-face context using an ANOVA including age as a between-subject factor. A 2 (sound type: speech vs. non-speech) \times 2 (age) mixed ANOVA revealed a main effect of sound type ($F(1,51) = 6.336$, $p = .015$, $\eta_p^2 = .111$), such that infants looked quicker to the Recipient following speech than non-speech, whereas the interaction between sound type and age was not significant ($F(1,51) = 2.322$, $p = .134$, $\eta_p^2 = .044$). In the back-to-back context, however, the same analysis did not reveal any significant differences between sound types ($F(1,51) = 1.388$, $p = .244$, $\eta_p^2 = .027$), suggesting that the effect of speech in decreasing latency to look at the Recipient was selective to the context where the two actors faced each other. To test directly for the effect of sound type in the face-to-face context for each age group, separate t-tests were performed. These revealed that the 24-month-olds looked quicker to the Recipient following speech ($t(25) = -2.966$, $p = .007$), whereas the 12-month-olds did not ($t(26) = -.681$, $p = .502$).

Additional analyses were performed on the latency measure where trials in which the infants did not look at the Recipient were excluded. The measure was square root transformed to a normal distribution. The omnibus ANOVA revealed a marginal main effect of orientation, with quicker looking to the Recipient in the face-to-face context ($F(1,44) = 3.904$, $p = .054$, $\eta_p^2 = .082$). For the face-to-face context, a 2 (sound type) \times 2 (age) mixed ANOVA revealed a marginally significant interaction between age and sound type ($F(1,50) = 3.685$, $p = .061$, $\eta_p^2 = .069$). Analyses of simple effects show that the 24-month-olds, but not the 12-month-olds, shifted their gaze significantly quicker to the Recipient following speech than non-speech (respectively, $F(1,50) = 5.91$, $p = .019$, $\eta_p^2 = .106$; $F(1,50) = .08$, $p = .778$, $\eta_p^2 = .002$). In the back-to-back context, however, the same analysis did not reveal any significant differences between sound type or age (respectively, $F(1,45) = .129$, $p = .721$, $\eta_p^2 = .003$; $F(1,45) = .973$, $p = .329$, $\eta_p^2 = .021$). Thus, when trials without looks to the Recipient are excluded, the 24-month-olds were found to look quicker to the Recipient following speech in the face-to-face context.

Recall that our main analysis included only infants who fixated the Emitter in two or more trials per condition. We reanalyzed the data using a less conservative exclusion criterion, where all infants who fixated the Emitter in at least a single trial per condition were included. Results were similar to the main analysis with the more conservative inclusion criteria. A 2 (sound type) \times 2 (age) mixed ANOVA's for the face-to-face context revealed a main effect of sound type ($F(1,72) = 4.890$, $p = .030$, $\eta_p^2 = .064$), but also an interaction between sound type and age, ($F(1,72) = 4.869$, $p = .031$, $\eta_p^2 = .063$), with simple effects again revealing that 24-month-olds looked quicker to the Recipient following speech than non-speech ($F(1,72) = 9.03$, $p = .004$, $\eta_p^2 = .111$), whereas 12-month-olds did not ($F(1,72) = 2.322$, $p = .997$, $\eta_p^2 = .000$). For the back-to-back context, no main effects or interactions were found.

Taken together, the latency analyses all reveal that in the face-to-face context, the 24-month-olds tended to look more quickly to the Recipient following speech than non-speech, whereas the 12-month-olds did not seem to distinguish between the sound types.

2.3. Looking time differences

Our second main measure was looking time differences. We analyzed the difference score for infants' looking time at the Recipient and the Emitter (looking time at Recipient – looking time at Emitter). For the sake of clarity, in the following description and in Experiment 2, higher difference scores (longer looking to the Recipient relative to the Emitter) are described as longer looking to the Recipient. Fig. 3 displays the means across conditions and age groups. The initial omnibus 2 (sound type: speech vs. non-speech) \times 2 (orientation: face-to-face vs. back-to-back) \times 2 (age: 12-months vs. 24-months) mixed ANOVA revealed a main effect of orientation, with longer looking at the Recipient in the face-to-face than in the back-to-back context ($F(1,51) = 7.639$, $p = .008$, $\eta_p^2 = .130$), and this effect interacted with sound type ($F(1,51) = 15.059$, $p < .001$, $\eta_p^2 = .228$). In addition, orientation interacted with age ($F(1,51) = 5.992$, $p = .018$, $\eta_p^2 = .105$). To directly address our main questions, we compared difference scores for speech vs. non-speech separately for each orientation as we did for the latency

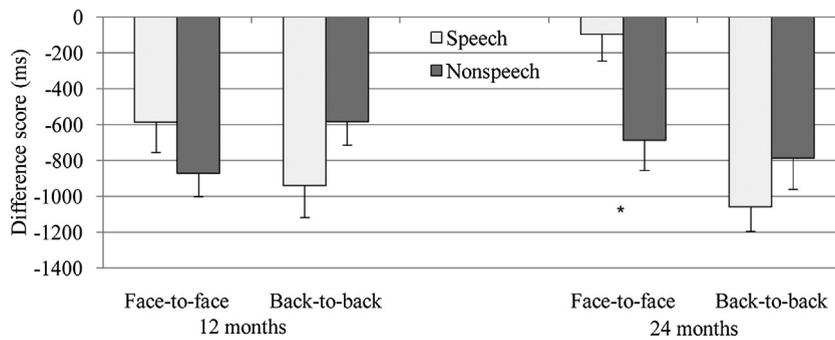


Fig. 3. Average difference scores of looking at the Recipient relative to the Emitter in Experiment 1, for all four conditions and both age groups. Difference scores were calculated by subtracting the looking time to the Emitter from the looking time to the Recipient during the post-sound phase. Asterisks indicate a significant difference ($p < .05$).

measure. The ANOVA's revealed that infants looked longer at the Recipient following speech than non-speech in the face-to-face context ($F(1,51) = 9.014, p = .004, \eta_p^2 = .150$), whereas the interaction between sound type and age was not significant ($F(1,51) = 1.101, p = .299$). In addition, there was a significant effect for age in the face-to-face context, revealing that the 24-month-olds looked overall longer at the Recipient in the face-to-face context than the 12-month-olds ($F(1,51) = 4.120, p = .048, \eta_p^2 = .075$). In the back-to-back context, however, infants looked less at the Recipient following speech than non-speech ($F(1,51) = 4.258, p = .044, \eta_p^2 = .077$). To test directly for the effect of sound type in the face-to-face context for each age group, separate t-tests were performed. These revealed that the 24-month-olds looked longer at the Recipient following speech than non-speech ($t(25) = 2.425, p = .023$), and that the 12-month-olds had a marginally significant tendency to do the same ($t(26) = 1.736, p = .094$).

As was done for the latency measure, 2 (sound type) \times 2 (age) mixed ANOVA's for the face-to-face and back-to-back contexts were performed on all infants who fixated the Emitter in at least one trial – a lower inclusion criteria than in the main analysis. Again, these results closely mirror our main analysis with the more conservative inclusion criteria. For the face-to-face context, the ANOVA revealed a main effect of sound type ($F(1,72) = 8.055, p = .006, \eta_p^2 = .101$), whereas the interaction between sound type and age was not significant ($F(1,72) = 2.370, p = .128, \eta_p^2 = .032$). For the back-to-back context, no main effects or interactions were found. Finally, to examine infants' attention to the overall visual display, we measured the proportion of time they spent looking at the two areas of interest, out of the total time they looked at the screen. On average, the 12-month-olds spent 83.3% of the time looking at the actors ($SD = 8\%$) and the 24-month-olds spent 82.4% of the time looking at the actors ($SD = 13\%$).

To summarize, our main analyses revealed that when the sounds were emitted in a face-to-face context, 24-month-olds, but not 12-month-olds, were quicker to shift their gaze to the Recipient following speech than non-speech. Similarly, 24-month-olds looked longer at the Recipient following speech than non-speech in the face-to-face context, and while 12-month-olds also tended to look longer following speech, this effect did not reach significance. In the back-to-back context, however, infants did not look quicker or longer to the Recipient following speech than non-speech, but instead looked at her for a shorter duration following speech. For both measures, additional analyses using less conservative inclusion criteria revealed highly similar findings.

Since the measures of gaze shift latency and looking time to the Recipient are influenced by how long infants look at the Emitter following her sound, at first glance it seems possible that the findings reflect not differing expectations toward the Recipient, but differing interests in the Emitter's sound. On this account, infants' longer or quicker looking to the Recipient following speech are due not to a stronger expectation that the Recipient will respond, but to a greater interest in – and thus longer looking to – the Emitter when she emits a non-speech sound. We believe this is an unlikely account for two reasons. First, if infants are more interested in the non-speech sounds, they should look longer at the Emitter following non-speech than speech, irrespective of her orientation. However, the results show that the effect was found only in the face-to-face context. In the back-to-back context, this effect either disappeared or reversed, with infants looking longer at the Recipient following non-speech (the equivalent of looking longer at the Emitter following speech). Second, research shows that infants look longer at human faces when they are accompanied by speech sounds than when they are accompanied by human non-speech vocalizations (Vouloumanos, Druhen, Hauser, & Huizink, 2009), suggesting that, if anything, infants would be expected to focus on the Emitter longer after speech as compared to non-speech. Thus, a general preference to look at the Emitter following non-speech is unlikely to account for our findings.

The results reveal that 24-month-old infants look quicker and longer to a third-party listener following speech compared to non-speech sounds, suggesting a stronger expectation of a response to speech. Their expectations are restricted to a setting where speech is uttered in a typical face-to-face conversational context. The 12-month-olds, on the other hand, showed little evidence of expecting a response to speech. Possibly, the 24-month-olds outperformed the younger age group because they have had more experience as both participants in and observers of conversations.

However, given the high attrition rate for the 12-month-olds, and the high number of them who failed to look at the Emitter during her sound, it is possible that the stimuli were simply not engaging enough to sustain their attention, and that they had difficulties realizing which actor had made the sound. Further, the within-subject design with four different conditions and 24 trials may have been too challenging for 12-month-olds' visual attention. Thus, in a second experiment, we made several changes to the paradigm to make the exchanges more realistic and easier to follow than the somewhat artificial exchanges in Experiment 1, and simplified the design to minimize the need to exclude data. The extent of these changes means that Experiment 2 was not a directly comparable variation on Experiment 1, but was designed as a simplified and more natural example of a verbal exchange that younger infants may find more engaging and easier to process. First, to make the social context more natural, the actors smiled and greeted each other before the exchange. The greeting made it clear that both actors were potentially willing to move and communicate. Second, in order to help infants identify the Emitter as the source of the sound, we attracted their attention to the Emitter by having her inhale visibly before emitting her sound. Finally, to reduce the number of trials excluded due to general inattentiveness, we presented fewer trials, using a between-subjects design. The between-subjects design also allowed us to compare infants' performance over trials, a comparison that was not possible in Experiment 1 as potential carryover effects restricted usable trials for such an analysis to the first block, which also led to a low number of infants for the comparison between conditions. Given the possibility that infants learn to anticipate responses through the contingency of conversations, it is important to examine whether they expect a response in the first trials, and how their expectations develop when they repeatedly view utterances that provoke no response.

Apart from improving the paradigm, Experiment 2 was also designed as a more rigorous test of infants' expectations. As gaze direction and body orientation of the actors in Experiment 1 matched, the older infants' looking in the face-to-face context may have been influenced by a tendency to follow the gaze direction of the Emitter after she uttered speech. Infants follow others' gaze direction from early in life (D'Entremont, Hains, & Muir, 1997; Scaife & Bruner, 1975), not only that of their communicative partners, but also that of people depicted in static images (von Hofsten, Dahlström, & Fredriksson, 2005). Although the specific role of speech in gaze-following has not been explored, infants show understanding of the referential nature of both words (Gliga & Csibra, 2009) and gaze (Csibra & Volein, 2008; Moll & Tomasello, 2004; Phillips et al., 2002), and may attend to the object of a speaker's gaze in search of a referent. Thus, it is possible that infants look quicker to and longer at a listener following face-to-face speech (than following face-to-face non-speech) because speech selectively prompts them to follow the speaker's gaze direction to the listener. To address this issue, in Experiment 2 we presented exchanges where the actors were facing each other in terms of body orientation, but were not in mutual gaze during the sound and instead looked at a piece of paper each of them held in their hand. Since the Emitter was not looking at the Recipient but at her piece of paper, infants' looking to the Recipient could not be influenced by the Emitter's gaze direction. Finally, Experiment 2 was specifically designed to probe and corroborate our findings with 12-month-olds, in order to explore further whether infants have expectations about speech already at the early stages of language acquisition, when they have little linguistic experience themselves, but have observed others in conversation and have participated in conversational turn-taking through proto-conversations.

3. Experiment 2

Experiment 2 examined 12-month-olds' expectations about third-party interactions further by employing more natural and engaging exchanges in which actors faced each other, but with averted gaze, while they either uttered speech or made a non-speech sound. As in Experiment 1, the Recipient never responded to the sounds, preventing infants from forming expectations based on previous contingent responses in the stimuli. Since we were especially interested in the performance of the younger age group, the second experiment included only 12-month-olds. The between-subjects design allowed us to examine changes in infants' expectations over time, testing two additional hypotheses. If infants' expectations regarding speech are present a priori, they should be seen already in the first half of the experiment, and possibly even in the first trial. Further, if contingent interactions as employed in previous paradigms (Augusti et al., 2010; Bakker et al., 2010; von Hofsten et al., 2009) facilitated anticipations, their absence in the current paradigm might lead to a decrease in expectations.

3.1. Method

3.1.1. Participants

Thirty-seven 12-month-olds (22 boys, 15 girls; mean age = 12;22, range = 12;00 to 12;30) participated, with 18 randomly assigned to the speech condition. Counterbalancing was based on 16 infants per condition, but in case of dropouts or cancellations, additional infants had been scheduled and these were included to increase power. Three additional infants participated but were not included in the final sample: two due to failing to look at the Emitter in at least two trials (see Data Reduction) and one due to not finishing the experiment as a result of fussiness. Infants were recruited as in Experiment 1.

3.1.2. Apparatus

Unlike in Experiment 1, infants' eye movements were measured with a Tobii T120 remote eye-tracker, using a sampling rate of 60 Hz. The eye-tracker has an accuracy of 0.5°, precision of 1°, and allows head-movements of up to 44 cm

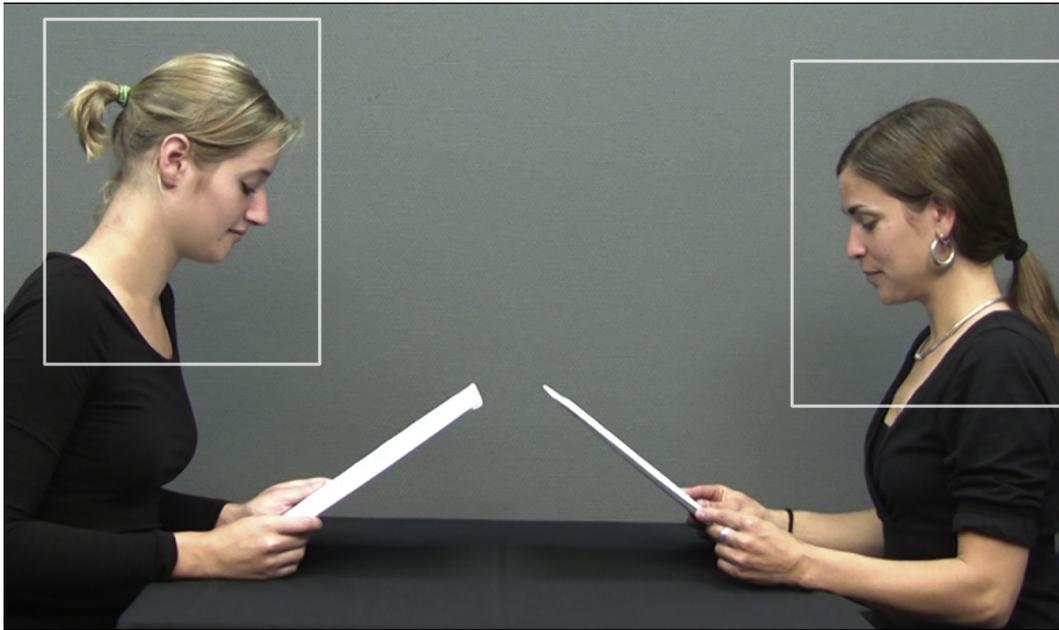


Fig. 4. A frame from the post-sound phase of a video from Experiment 2, where the Emitter is on the right side. The rectangles define the areas of interest.

horizontally, 22 cm vertically, and 30 cm in depth. The eye tracker is integrated with a 17" TFT display with a native resolution of 1280×1024 pixels.

3.1.3. Stimuli

Infants were presented with eight videos (29.5×17.33 visual degrees), which lasted for 1.5 min total. Each video began with the actors waving and greeting each other. Importantly, the greetings were completely synchronous in order to avoid the impression of contingent reactivity. Then each picked up a piece of paper and looked at it. The Emitter (shown on the right side in Fig. 4) then inhaled visibly and audibly (2 s) before she emitted a sound (mean duration = 1.51 s). After the sound, both actors were inactive (4 s) (see Fig. 4). In the speech condition, the Emitter uttered a question or a statement, and in the non-speech condition, she either cleared her throat or yawned (see Table 2). For both conditions, the two sound examples were presented in alternation, and each pair was presented twice with the Emitter on the right side of the screen and twice with the Emitter on the left side (video was flipped horizontally), also in alternation. Example type of the first video and Emitter position of the first pair was counterbalanced across infants.

3.1.4. Procedure

Infants were seated in a safety car seat that was placed in their parents' lap so that the infants' eyes were approximately 64 cm from the monitor. They saw a short animation (a colorful looming ball accompanied by sound with a duration of 4 s) between trials 4 and 5 to reengage their attention in case it had drifted away from the screen. As the greeting at the start of each video was highly salient, and the Emitter drew attention to herself before she made the sound, no attention getters were displayed between the other trials. In other respects, the procedure was identical to that used in Experiment 1.

3.1.5. Data reduction

As in Experiment 1, rectangular AOI's were created around each of the actors' faces (7.77×9.75 visual degrees each; see Fig. 4) covering an area approximately 30 pixels (0.71 visual degrees) wider and higher than the actors. Unfiltered gaze points registered in these areas were used to calculate the measures of gaze shift latency and difference score. The sound phase and the post-sound phase were selected as in Experiment 1, and the post-sound phase had the same duration in both conditions (3.8 s). As in Experiment 1, for the latency analysis, trials in which infants did not look at the Recipient were

Table 2
Characteristics of the Speech and the Non-speech Sounds in Experiment 2.

Speech sound utterances	Non-speech sounds
"Is de kat al buiten?" (Is the cat already outside?)	Throat clearing
"Het is lekker weer" (The weather is nice)	Vocal sigh
Mean duration = 1360 ms, SD = 56.57	Mean duration = 1660 ms, SD = 84.85

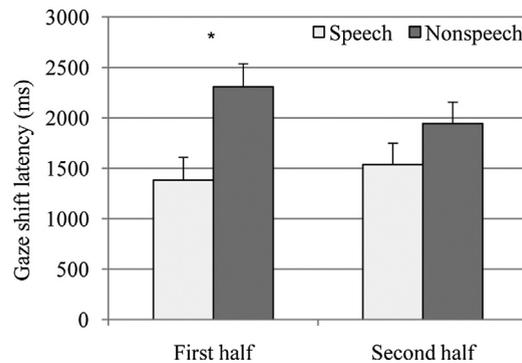


Fig. 5. Average gaze shift latency in the speech and non-speech conditions of Experiment 2, for the first and the second half of the trials. Asterisks indicate a significant difference ($p = .05$).

given a maximum score (3800). For both dependent measures, trials in which the infants did not fixate the Emitter during the sound were excluded. On average, 2.27 trials out of 8 per infant were excluded for this reason, but the number of trials excluded did not differ between conditions ($t(35) = .341$, $p = .735$). Infants who did not fixate the Emitter in at least two trials were excluded from the analysis ($n = 2$; both in the non-speech condition). To measure fixations to the Emitter during speech we used the Tobii Studio fixation filter, with the default 35 pixel velocity and distance thresholds. This newer and more accurate fixation algorithm is slightly different than the one used in Experiment 1, with a minimum fixation length of five data points (83 ms rather than 100 ms). We re-analyzed the current data with the fixation filter from Experiment 1. When using the filter used in Experiment 1, more trials had to be excluded due to the lower sensitivity of the filter (36 trials in total), but the main findings remained the same (see Section 3.2).

3.2. Results and discussion

We conducted ANOVAs on both dependent measures with condition as a between subjects variable and first vs. second trial block of the experiment as a repeated measure to examine performance over trials. The gaze shift latency measure revealed that infants looked quicker to the Recipient following speech than non-speech ($F(1,34) = 5.896$, $p = .021$, $\eta_p^2 = .148$). A marginally significant interaction was found between sound type and block ($F(1,34) = 3.197$, $p = .083$, $\eta_p^2 = .086$). To address our question of learning over trials we conducted a simple effects analysis which revealed that infants looked quicker to the Recipient following speech than non-speech in the first block of the experiment ($F(1,34) = 8.34$, $p = .007$, $\eta_p^2 = .197$), but not in the second block ($F(1,34) = 1.84$, $p = .184$, $\eta_p^2 = .051$; see Fig. 5). An analysis on the first trial (in which infants fixated the Emitter) similarly revealed that infants in the speech condition were quicker to look to the Recipient than infants in the non-speech condition ($t(35) = -2.770$, $p = .009$; respectively, $M = 1453$, $SD = 1224$, $M = 2590$, $SE = 1270$).

Several additional analyses were performed on the latency measure. First, a 2 (sound type) \times 2 (block) ANOVA was performed where trials without a look to the Recipient were excluded instead of given a maximum latency. This analysis revealed a main effect of sound type ($F(1,29) = 4.659$, $p = .039$). A first trial analysis on the same data revealed a marginally significant effect of sound type, such that infants in the speech condition looked quicker to the Recipient in the first trial than infants in the non-speech condition, $t(34) = -1.922$, $p = .063$. Finally, these analyses were repeated after applying the fixation filter from Experiment 1. The 2×2 ANOVA revealed a main effect of sound type ($F(1,31) = 7.709$, $p = .009$, $\eta_p^2 = .199$), such that infants in the speech condition were overall quicker to look at the Recipient, and the first trial analysis revealed that they did so already on the first trial, $t(32) = -2.983$, $p = .005$.

An ANOVA for the difference score measure, using the same factors as the analysis for latency, revealed a similar pattern of looking as the 24-month-olds showed in Experiment 1, with longer looking to the Recipient than the Emitter. However, these differences were not statistically significant ($F(1,34) = 1.240$, $p = .273$, $\eta_p^2 = .035$), nor were there other significant main effects or interactions (p 's $> .10$; see Fig. 6). Finally, to examine the infants' overall attention to the scene, we created additional Areas of Interest around the actors' hands and the paper they were holding. On average, out of the time infants looked at the screen during the post-sound phase, they looked at the Emitter 37% of the time ($SD = 13\%$), the Recipient 29% of the time ($SD = 12\%$), the Emitter's hand holding the paper 6% of the time ($SD = 7\%$), the Recipients hand and paper 8% of the time ($SD = 5\%$), and spent 20% of the time looking at other areas ($SD = 16\%$).

It is theoretically possible that infants did not detect the Emitter's downward gaze direction, but instead assumed from the body orientation of the two actors that the Emitter's attention was directed to the Recipient. However, based on the reviewed literature of infants' understanding of gaze direction we find this unlikely. Further, after the Emitter has greeted the Recipient, and before she makes the sound, she shifts her gaze down from the Recipient toward the paper, and then picks up the paper and tilts it upward. These events make the Emitter's downward gaze direction highly salient and presumably easy for infants to notice, especially given that infants reliably follow gaze direction (e.g. Senju & Csibra, 2008), even when gaze does not change direction (von Hofsten et al., 2005).

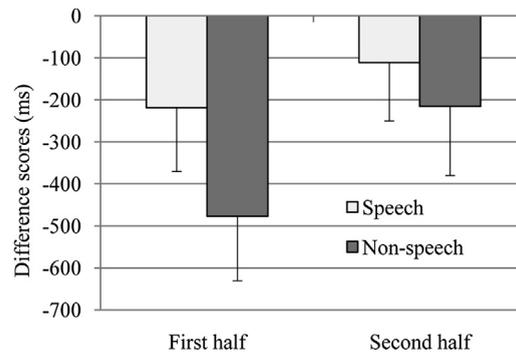


Fig. 6. Average difference scores of looking at the Recipient relative to the Emitter in the speech and non-speech conditions of Experiment.

To summarize, infants in the speech condition were quicker to shift their gaze to the Recipient than infants in the non-speech condition, suggesting that they had stronger expectations for her to respond to speech. As the two actors were not in mutual gaze during the sound, but instead gazed down to their papers while facing each other in body orientation, the differences found between conditions could not have been due to infants' selectively following the gaze-direction of the Emitter after she spoke. Interestingly, the latency measure indicates that infants' differential expectations of a response to speech over non-speech ceased in the second half of the experiment. Unlike Experiment 1, no significant differences were found for the measure of looking time, although the means were in the predicted direction. The lack of this effect may be due the fact that infants distributed their attention more evenly to the more complex visual display presented in Experiment 2 and spent on average roughly third of the time looking at other areas, compared to only around 17% in Experiment 1. Further, as both actors were more active in Experiment 2 (greeting each other and picking up and looking at papers), infants may have equally expected further actions from both of them.

4. General discussion

We found that 12- and 24-month-old infants looked quicker to a third-party listener following speech compared to non-speech sounds, suggesting a stronger expectation of a response to speech. In Experiment 1, only the 24-month-olds showed evidence of expecting a response, whereas 12-month-olds did so when presented with the more natural and animated exchanges in Experiment 2. In addition, the 24-month-olds' were found to have stronger expectations toward the listener when the speaker and listener faced each other than when they were back-to-back. The findings appear to indicate that infants distinguish speech from other natural human sounds as having the potential to produce a response from another person, at least when the orientation of the speaker and listener is typical for a communicative interaction. In addition, these looking patterns could not have resulted from a tendency to follow the gaze of the Emitter to the Recipient, since in Experiment 2, where the two actors averted their gaze, 12-month-olds showed evidence of having a stronger expectation of a response to speech. Thus, the current findings suggest that infants in the early stages of language acquisition not only monitor ongoing third-party conversations (Augusti et al., 2010; Bakker et al., 2010; von Hofsten et al., 2009), but also expect third-party communicative acts to provoke a response. A recent looking time study showed that 12-month-olds understand that speech, as opposed to a cough or an emotional vocalization, can function to transfer information about an object to an addressee in a third-party interaction (Martin et al., 2012). The current finding complements and extends this previous study with online-processing measures by showing that infants also seem to actively anticipate that third-party addressees will respond to speech, but not to natural non-speech sounds.

At first glance, our findings may seem at odds with those from a previous study that included comparisons between infants' looking at a normal conversation and the same conversation but with the speech replaced by mechanical sounds (Bakker et al., 2010). No differences were found in how often 1-year-olds or 3-year-olds disengage from the current to the prospective speaker, nor how long they looked at the scene. However, since humans do not emit mechanical sounds, infants in that condition may have perceived the back-and-forth sounds between the two people looking at each other as a form of communication (e.g. Johnson, Slaughter, & Carey, 1998). In contrast, the current study used natural non-speech sounds familiar to infants, which may have helped them distinguish non-communicative from communicative sounds. Further, in contrast to previous studies (Augusti et al., 2010; Bakker et al., 2010; von Hofsten et al., 2009), the design of the current experiments did not involve contingent reactions, which, if contingent on infants' own behavior, can lead them to infer communication even from non-human, non-verbal agents (Deligianni, Senju, Gergely, & Csibra, 2011). Instead, both experiments controlled for contingency between the actors so that it could not influence infants' gaze patterns. Even the initial greeting in Experiment 2 was synchronous and thus did not demonstrate contingent turn taking.

The relatively minimalistic design of the current experiments and the nature of the dependent measures would seem to leave the results open to alternative accounts. Since infants' looking toward the Recipient is directly related to their looking at the Emitter, perhaps infants' longer and quicker looking at the Recipient following speech is the result of infants

looking longer at the Emitter when she produces non-speech, possibly because these events are more unexpected. However, such a preference would have led to longer looking at the Emitter irrespective of the actors' orientation, whereas infants in Experiment 1 only showed the effects in the face-to-face context. Moreover, previous research demonstrates that infants show the opposite bias – they look longer at human faces when they are accompanied by speech as compared to non-speech (Vouloumanos et al., 2009). Another account might posit that infants' looking patterns reflect not their expectations toward an addressee but a propensity to look in the same general direction that a speaker is facing or looking at. Such a propensity, one could argue, might in turn be the result of speech triggering an automatic gaze-following response, or of an understanding of speech as an action not directed specifically to another person, but a non-specific target. However, the design of Experiment 2, where the Emitter faced the Recipient but averted her gaze downward during the exchange, ensured that infants' looking to the Recipient was not influenced by the Emitter's specific gaze direction. In addition, previous work similarly indicates that gaze-following does not determine the extent to which infants monitor speakers in a conversation (Augusti et al., 2010). Finally, while we did not compare infants' looking to a human addressee with their looking to a non-human target, looking time studies suggest that at least by 10 months, if not earlier (Legerstee, Barna, & DiAdamo, 2000), infants expect people to converse with and look at other people, rather than inanimate objects (Beier & Spelke, 2012).

One remaining question is why the 12-month-olds showed evidence of stronger response expectations following speech in Experiment 2, but not in Experiment 1. We did not design Experiment 2 to make a direct statistical comparison to Experiment 1, but in Experiment 2 we made various modifications to the paradigm from Experiment 1 that we believe succeeded in facilitating infants' ability to process and follow the exchange. For example, infants were much more likely to look at the Emitter during the sound in Experiment 2 – where the Emitter drew attention to herself before the sound by inhaling – than in Experiment 1. The greeting that preceded the sound in Experiment 2 may also have helped infants recognize that the actors were generally social and communicative, or made it clearer that they were socially engaged with each other. However, as these procedures were identical in the speech and the non-speech conditions, they cannot account for the differences we found between conditions.

Evidence for infants' increased expectations for the Recipient to respond to speech in Experiment 1 was only found when the actors were facing each other. Previous research shows that infants as young as 6 months are sensitive to whether conversational partners are face-to-face in mutual gaze, or back-to-back (Augusti et al., 2010), and at 10 months they also distinguish between these contexts for static, silent images (Beier & Spelke, 2012). The findings from Experiment 1 indicate that at least by 2 years of age, infants also see the face-to-face orientation with mutual gaze as a setting in which speech is likely to provoke a response. This may be because they understand the face-to-face orientation as a typical communicative context, having learned that speakers typically face their addressees. As the 12-month-olds in Experiment 1 did not have increased expectations in either context, and Experiment 2 did not manipulate the gaze direction or body orientation of the actors, it is less clear how body orientation or mutual gaze affects expectations at this age. Nevertheless, 12-month-olds in Experiment 2 showed evidence of having a stronger expectation of a response to speech when the actors were face-to-face but with their gaze averted, suggesting that mutual gaze during speech is not a prerequisite for them to expect a response from a third-party listener, though face-to-face body orientation may be. Infants' perception of third-party communication may also be influenced by the action context within which communication takes place. One interesting avenue for further research is examining if infants' understanding of and expectations about speech are also influenced by whether or not the speaker and the addressee are simultaneously engaged in a joint activity.

Although infants in the experiments looked earlier toward the person addressed with speech, the latency of their switch toward the addressee was considerably longer than the average interval of approximately 100 ms between turns in Dutch conversations (Stivers et al., 2009). The fact that the infants in the current study were looking not at an ongoing conversation, but a speaker and a nonresponsive addressee, is likely to have contributed to the relatively long latency. However, previous studies have found that even when watching an ongoing conversation with typical turn intervals, it is only by three years of age that infants begin to consistently switch their gaze from a current to an upcoming speaker (Augusti et al., 2010; Bakker et al., 2010; Keitel, Prinz, Friederici, von Hofsten, & Daum, 2013). This suggests that while infants anticipate a response to speech from early on, successful prediction of the timing of turn-taking is a more challenging skill that takes longer to develop.

It is unclear how much of the semantic content of the utterances infants understood, and whether this influenced their expectations. It cannot be ruled out that the 24-month-olds outperformed the 12-month-olds in Experiment 1 partly because they understood more language. Infants' growing linguistic experience is certainly likely to explain their increased general interest in others' conversations and the flexibility with which they monitor speakers, demonstrated in previous eye tracking studies (Bakker et al., 2010; von Hofsten et al., 2009). However, a recent study showed that infants at 12 months expect even unfamiliar (nonsense) words to be able to convey information to a third party (e.g. Martin et al., 2012), indicating that infants' expectations toward addressees may depend only on recognizing the form of speech, but not understanding its semantic content.

A further question is how infants develop expectations about third-party communicative actions. Their expectations are likely partly based on observing and overhearing third-party interactions. While our stimuli presented no contingent reactions that infants could learn from, it is likely that their expectations about speech stem at least in part from previous experience with others' contingent turn-taking. Infants' expectations may also be informed by their own experience in communication. The current results, along with previous findings (Martin et al., 2012), suggest that infants already have expectations toward third-party addressees of speech at an age where they have very minimal experience with

communicating through speech themselves. However, by their first birthday, infants have nonetheless typically had plenty of communicative experience, where they engage in a coordinated fashion with others' speech through various actions, including speech-like babbling (e.g. Feldstein et al., 1993; Jaffe, Beebe, Feldstein, Crown, & Jasnow, 2001). Participation in such proto-conversations may give rise to expectations about the role of speech in others' interactions, perhaps based on an understanding of the basic communicative function of speech. It will be important to examine whether infants' understanding of communicative actions addressed to them is prerequisite for their understanding of third-party communication, or when these two integrate and bring infants to a more abstract understanding of the communicative act (see also Moore, 2007).

The current study adds to recent findings on infants' action understanding. Beyond infants' expectations about others' object-directed actions, the study shows that at the time they are only beginning to communicate through speech infants expect others' speech actions to be directed at persons. This finding suggests that infants understand communication from an observer's perspective as a shared activity with distributed roles, which goes beyond infants' second-person understanding evident in their own interactions. This skill is valuable because attending to the response of a third-party addressee enables social learning not only about individual object-directed goals but also about social goals of acting together (see also Fawcett & Liszkowski, 2012), and may even be a major route of learning in communities where infants and children are less frequently directly addressed.

Acknowledgements

We thank Merel Fokkema, Dirkje van der Aa, and other actors for their assistance, as well as all the participants and their families.

References

- Augusti, E. M., Melinder, A., & Gredebäck, G. (2010). Look who's talking: Pre-verbal infants' perception of face-to-face and back-to-back social interactions. *Frontiers in Psychology: 1*, 161.
- Bakker, M., Kochukhova, O., & von Hofsten, C. (2010). Development of social perception: A conversation study of 6-, 12- and 36-month-old children. *Infant Behavior and Development: 34*, (2), 363–370.
- Bateson, M. C. (1975). Mother-infant exchanges. The epigenesis of conversational interactions. *Annals of the New York Academy of Sciences: 263*, (1), 101–113.
- Becker, W. (1972). The control of eye movements in the saccadic system. *Bibliotheca Ophthalmologica: 82*, 233–243.
- Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science: 8*, (6), 492–499.
- Beier, J. S., & Spelke, E. S. (2012). Infants' developing understanding of social gaze. *Child Development: 83*, 486–496.
- Brown, P. (1998). Conversational structure and language acquisition: The role of repetition in Tzeltal. *Journal of Linguistic Anthropology: 8*, (2), 197–221.
- Canfield, R. L., & Haith, M. M. (1991). Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology: 27*, (2), 198–208.
- Canfield, R. L., Smith, E. G., Brezsnayak, M. P., & Snow, K. L. (1997). Information processing through the first year of life: A longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development: 62*. [2, serial no. 250].
- Csibra, G., & Volein, A. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology: 26*, 1–11.
- D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development: 20*, 569–572.
- Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental Psychology: 47*, (6), 1499–1503.
- Doi, H., Tagawa, M., & Shinohara, K. (2010). Gaze direction modulates the disengagement of attention from facial expression in 10-month-olds. *Emotion: 10*, (2), 278–282.
- Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience: 9*, 878–879.
- Fawcett, C., & Liszkowski, U. (2012). Observation and initiation of joint action in infants. *Child Development: 83*, (2), 434–441.
- Feldstein, S., Jaffe, J., Beebe, B., Crown, C. L., Jasnow, M., Fox, H., et al. (1993). Coordinated interpersonal timing in adult infant vocal interactions – a cross site replication. *Infant Behavior and Development: 16*, (4), 455–470.
- Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). Amsterdam: John Benjamins.
- Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy: 9*, (3), 327–339.
- Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science: 20*, 347–353.
- Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). One-year-olds' understanding of nonverbal gestures directed to a third person. *Cognitive Development: 24*, (1), 23–33.
- Gredebäck, G., & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition: 114*, (2), 197–206.
- Heath, S. B. (1983). *Ways with words: Language, life and work in communities and classrooms*. Cambridge University Press.
- Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology: 46*, (2), 446–454.
- Hunnius, S., de Wit, T. C. J., Vriens, S., & von Hofsten, C. (2011). Facing threat. Infants' and adults' visual scanning of faces with neutral, happy, sad, angry, and fearful emotional expressions. *Cognition and Emotion: 25*, 193–205.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development: 66*, (2), i–viii, 1–132.
- Johnson, S., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. *Developmental Science: 1*, 233–238.
- Keitel, A., Prinz, W., Friederici, A. D., Hofsten, C. V., & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology: 116*, (2), 264–277.
- Legerstee, M., Barna, J., & DiAdamo, C. (2000). Precursors to the development of intention at 6 months: Understanding people and their actions. *Developmental Psychology: 36*, (5), 627–634.
- Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway, & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56–73). Cambridge, UK: Cambridge University Press.

- Martin, A., Onishi, K. H., & Vouloumanos, A. (2012). Understanding the abstract role of speech in communication at 12 months. *Cognition*: 123., 50–60.
- Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*: 7., F1–F9.
- Moore, C. (2007). Understanding self and others in the second year. In C. Brownell, & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 43–65). New York: Guilford Press.
- Nielsen, M. (2009). 12-month-olds produce others' intended but unfulfilled acts. *Infancy*: 14., (3), 377–389.
- Nurmsoo, E., Shiri, E., & Hood, B. M. (2012). Best friends: Children use mutual gaze to identify friendships in others. *Developmental Science*: 15., (3), 417–425.
- Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization: Three developmental stories and their implications. In Richard Schweder, & Robert A. Levine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 276–320). Cambridge: Cambridge University Press.
- Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*: 85., 53–78.
- Reddy, V. (2008). *How infants know minds*. Cambridge, MA: Harvard University Press.
- Repacholi, B. M., & Meltzoff, A. N. (2007). Emotional eavesdropping: Infants selectively respond to indirect emotional signals. *Child Development*: 78., (2), 503–521.
- Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*: 253., 265–266.
- Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*: 18., (9), 668–671.
- Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*: 96., (1), B1–B11.
- Southgate, V., & Csibra, G. (2009). Inferring the outcome of an ongoing novel action at 13 months. *Developmental Psychology*: 45., (6), 1794–1798.
- Southgate, V., Johnson, M. H., El Karoui, I., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological Science*: 21., (3), 355–359.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Science*: 106., (26), 10587–10592.
- Thayer, S., & Schiff, W. (1974). Observer judgment of social interaction: Eye contact and relationship inferences. *Journal of Personality and Social Psychology*: 30., 110–114.
- Thorgrimsson, G. B., Fawcett, C., & Liszkowski, U. (2014). Infants' expectations about gestures and actions in third-party interactions. *Frontiers in Psychology*: 5., 321.
- Vouloumanos, A., Druhen, M. J., Hauser, M. D., & Huizink, A. T. (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*: 106., (44), 18867–18872.
- von Hofsten, C., Dahlström, E., & Fredriksson, Y. (2005). 12-month old infants' perception of attention direction in static video images. *Infancy*: 8., 217–231.
- von Hofsten, C., Uhlig, H., Adell, M., & Kochukhova, O. (2009). How children with autism look at events. *Research in Autism Spectrum Disorders*: 3., (2), 556–569.
- Wang, S., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. *Cognition*: 93., 167–198.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*: 69., 1–34.