Check for updates

Received: 13 September 2022 | Revised: 20 April 2023 | Accepted: 4 July 2023

DOI: 10.1111/desc.13436

Developmental Science WILEY

RESEARCH ARTICLE

Does the speaker's eye gaze facilitate infants' word segmentation from continuous speech? An ERP study

Melis Çetinçelik[1] | **Caroline F. Rowland**[1,2] | **Tineke M. Snijders**[1,2,3]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, Gelderland, The Netherlands

[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Gelderland, The Netherlands

[3]Cognitive Neuropsychology Department, Tilburg University, Tilburg, The Netherlands

Correspondence
Melis Çetinçelik, Max Planck Institute for Psycholinguistics, Nijmegen, Gelderland, The Netherlands.
Email: melis.cetincelik@mpi.nl

Funding information
Max Planck Society

Abstract

The environment in which infants learn language is multimodal and rich with social cues. Yet, the effects of such cues, such as eye contact, on early speech perception have not been closely examined. This study assessed the role of ostensive speech, signalled through the speaker's eye gaze direction, on infants' word segmentation abilities. A familiarisation-then-test paradigm was used while electroencephalography (EEG) was recorded. Ten-month-old Dutch-learning infants were familiarised with audio-visual stories in which a speaker recited four sentences with one repeated target word. The speaker addressed them either with direct or with averted gaze while speaking. In the test phase following each story, infants heard familiar and novel words presented via audio-only. Infants' familiarity with the words was assessed using event-related potentials (ERPs). As predicted, infants showed a negative-going ERP familiarity effect to the isolated familiarised words relative to the novel words over the left-frontal region of interest during the test phase. While the word familiarity effect did not differ as a function of the speaker's gaze over the left-frontal region of interest, there was also a (not predicted) positive-going early ERP familiarity effect over right fronto-central and central electrodes in the direct gaze condition only. This study provides electrophysiological evidence that infants can segment words from audio-visual speech, regardless of the ostensiveness of the speaker's communication. However, the speaker's gaze direction seems to influence the processing of familiar words.

KEYWORDS
audio-visual speech, ERP, infant EEG, language acquisition, word segmentation

Research Highlights

- We examined 10-month-old infants' ERP word familiarity response using audio-visual stories, in which a speaker addressed infants with direct or averted gaze while speaking.
- Ten-month-old infants can segment and recognise familiar words from audio-visual speech, indicated by their negative-going ERP response to familiar, relative to novel, words.

- This negative-going ERP word familiarity effect was present for isolated words over left-frontal electrodes regardless of whether the speaker offered eye contact while speaking.
- An additional positivity in response to familiar words was observed for direct gaze only, over right fronto-central and central electrodes.

1 INTRODUCTION

One of the first tasks infants have to address when building a vocabulary is to recognise word tokens by segmenting them from continuous speech, thus storing isolated word forms onto which they can map meaning. This is a challenging task, as most of the words that infants hear do not occur in isolation, but in continuous speech without cues that mark boundaries between individual words (Brent & Siskind, 2001; Kooijman et al., 2005; Van De Weijer, 1999).

Infants' ability to segment words from continuous speech is well-established. In a pioneering study using the Headturn Preference Paradigm (HPP), Jusczyk and Aslin (1995) demonstrated that 7.5-month-old infants showed a preference for (i.e. listened longer to) passages that contained words that they were first familiarised with in isolation, which means that they were able to segment previously heard word tokens from continuous speech. In another experiment, 7.5-month-olds recognised isolated words at test after being familiarised with sentences (Jusczyk & Aslin, 1995), pointing at the flexibility of their segmentation abilities. Several other behavioural studies have confirmed these findings in different languages, presenting robust evidence that 6–12-month-old infants are capable of segmenting words from a continuous speech stream (e.g. Houston et al., 2000; Jusczyk et al., 1999; Nazzi et al., 2005; Saffran et al., 1996; see Bergmann & Cristia, 2016 for a meta-analysis), although this might hold only when speech has exaggerated infant-directed properties, which is typical of American IDS (Fernald et al., 1989; Thiessen et al., 2005; see also Floccia et al., 2016).

Furthermore, neuroimaging studies focused on infants' event-related potentials (ERPs) reveal their rapid word recognition and segmentation abilities. These studies make use of the word familiarity ERP effect as an electrophysiological marker of word segmentation, by comparing ERPs time-locked to familiarised versus novel words in continuous speech, giving an online measure of word segmentation with high temporal precision (Kooijman et al., 2008). For example, Kooijman et al. (2005) familiarised 10-month-old infants with words in isolation, and then studied their ERPs to those familiarised and unfamiliar words as they occurred in continuous speech. In this study, the response in the familiarisation phase indexed only a word recognition response, as the word tokens were presented in isolation, but, in the test phase, the response required both word segmentation from continuous speech and recognition of the word as a familiar word form. In both the familiarisation and the test phase, 10-month-old infants showed a negative-going ERP response, although their responses differed in the latency and localisation of the familiarity effect in the two phases of the experiment. Infants had a very early, mainly left-lateralised, response to the repeated isolated word tokens starting at around 160 ms after word onset, indicating that they recognised the familiar word without even hearing the full word. The responses to the repeated words in continuous speech were observed slightly later and were more widely distributed across the brain, spanning frontal, fronto-central and fronto-temporal areas. Overall, this study demonstrated that, by 10 months, infants can recognise familiar words in continuous speech, as indicated by their negative-going word familiarity effect, with a more negative amplitude to the familiarised words compared to novel words.

This negative-going response, which is largely observed over left-frontal electrodes, has been reported in several electroencephalography (EEG) studies with infants using similar paradigms (Goyet et al., 2010; Kooijman et al., 2009, 2013), as well as paradigms that used continuous speech in both familiarisation and test (Junge et al., 2014) and those that use continuous speech in the familiarisation phase and single words in the test phase (Junge et al., 2012; Männel & Friederici, 2013). Some studies, however, have demonstrated a positive-going response, particularly for younger infants of 6–7 months of age (Kooijman et al., 2013; Männel & Friederici, 2013), or in experiments with more difficult stimuli with respect to acoustic prominence or speaking rate (Snijders et al., 2020). Hence, there seems to be a developmental shift with regards to the polarity of the ERP familiarity effect, with the negative ERP familiarity effect being considered a more mature response, possibly linked to the onset of more active lexical processing (Kidd et al., 2018; Männel & Friederici, 2013; Snijders et al., 2020).

However, most previous studies investigating word segmentation have focused only on unimodal auditory speech. In naturalistic conversation, speech is accompanied by a range of cues that facilitate language processing in general, and perhaps segmentation in particular. These cues are not only linguistic (e.g. prosodic cues in infant-directed speech, which facilitate speech processing and word segmentation; Thiessen et al., 2005), but are also multimodal (e.g. cues from the speaker's face), and may guide listeners to selectively attend to the relevant speech input.

The multimodal nature of speech potentially plays an important role in language acquisition, as infants learn language in multimodal communicative contexts, often with face-to-face communication with their caregivers. During face-to-face communication, many caregivers frequently use ostensive signals such as infant-directed speech and eye contact to direct infants' attention to the relevant input (Lavelli & Fogel, 2005; Soderstrom, 2007). Eye gaze stands out as an important cue in child–caregiver interactions, both in the form of object-directed gaze and infant-directed gaze (eye contact). It has been well-documented that infants are attentive to this type of gaze signalling in social settings: they are highly sensitive to eye gaze from birth (Farroni et al., 2006) and can follow others' gaze direction from 6 months of age (D'Entremont, 2000). Thus, eye gaze is a readily available and accessible cue in the environment, even for very young infants. The use of eye gaze cues by infants' communicative partners (usually, caregivers) in the form of object-directed and infant-directed gaze also plays a facilitatory role in infants' learning, and in particular, language development, such as infants using the adult's gaze direction to form word-object mappings (Gogate et al., 2006; Graham et al., 2007; Hollich et al., 2005; see Çetinçelik et al., 2021 for a review). Accordingly, early gaze-following and responding to joint attention abilities are positively correlated with later receptive and expressive vocabulary development (Brooks & Meltzoff, 2008; Morales, Mundy, Delgado, Yale, Messinger, et al., 2000; Morales, Mundy, Delgado, Yale, Neal, et al., 2000; see Çetinçelik et al., 2021 for a review).

Since infants prefer, and selectively attend to, faces with direct eye contact (Farroni et al., 2006), it might be the case that the information provided by people engaged in eye contact with infants is processed differently as well. Natural Pedagogy theory argues that during periods of mutual engagement, such as when accompanied by eye contact, information transfer between infants and adults is optimized (Csibra & Gergely, 2009). On this view, eye contact conveys to the child that the speaker intends to express a communicative act, which brings the child into a receptive state for upcoming information (Senju & Csibra, 2008), perhaps because the use of social cues in communication results in high-excitability oscillatory periods for optimal information encoding (Wass et al., 2020). Following this line of thought, we would, thus, expect eye contact to enhance learning in speech perception tasks such as word segmentation.

Even though it has been argued that word segmentation and statistical learning from speech occur incidentally (Saffran et al., 1997), there is, in fact, a role for attention. Adults' performance on word recognition from continuous speech has been reported to be significantly lower when they performed distractor tasks (Toro et al., 2005) and infants' speech perception has been shown to be hindered in the presence of distracting sounds (Polka et al., 2008), suggesting that attention to the speaker (and the speech input provided by the speaker) likely plays a role in learning from speech. Thus, there is good reason to predict a role for eye gaze, via attention, on infants' ability to segment words from the speech stream.

The goal of the present paper was to assess the role of ostensive speech, signalled through speaker's eye gaze direction, on infants' word segmentation abilities. While there is substantial literature on the relationship between eye gaze (in the form of gaze following) and vocabulary development, particularly word-object mappings, there is little evidence about the role of eye gaze on other language acquisition and processing tasks (Çetinçelik et al., 2021). In the domain of speech processing, there is some evidence that infants process spoken words differently when speech is accompanied by direct and averted eye gaze (Parise et al., 2011), and that their neural responses to speech may be enhanced when the speaker establishes eye contact with them when speaking (Lloyd-Fox et al., 2015). Similarly, adult listeners' speech and gesture processing may be facilitated when they are addressed with direct gaze by the speaker compared to an unaddressed recipient condition signalled through averted gaze (Holler et al., 2014). However, although these studies point to the possibility of speech being processed differently in the presence or absence of multimodal ostensive cues, they do not directly assess whether infants' processing of, and learning from, continuous speech is enhanced and whether they learn more from the speech input in the presence of direct gaze.

In this study, we investigated whether 10-month-old infants' learning from speech is enhanced when speech is presented in an ostensive manner, with direct eye contact. In particular, we tested infants' word segmentation from continuous speech as indexed by the word familiarity ERP effect, and how this effect differs when a speaker communicated with an infant using direct versus averted gaze. We focused on the left-frontal electrodes, as this region has been most consistently reported in infant EEG studies looking at word segmentation (e.g. Junge et al., 2012, 2014; Kidd et al., 2018; Kooijman et al., 2005, 2009, 2013; Männel & Friederici, 2013; Snijders et al., 2020). Importantly, we kept the prosodic features of speech constant across conditions, and only manipulated the speaker's gaze direction, so that any observed effects could be attributed to differential processing in the presence and absence of the added ostensive cue: eye gaze directed at or averted from the listener. Since our main aim was to investigate whether speaker's gaze affects learning, not whether it affects in-the-moment online processing, we first familiarised infants with the stimuli in the presence or absence of eye contact using audio-visual stimuli, and then assessed how they react to these prior familiarised stimuli in an audio-only test phase. However, we also report an additional analysis of the word familiarity ERP response in the familiarisation phase.

2 | METHODS

This study was pre-registered (https://aspredicted.org/blind.php?x=VSG_ZNB). Note that the pre-registration included an additional assessment of infants' language skills using the short Dutch versions of the MacArthur-Bates Communicative Development Inventory (N-CDI; Zink & Lejaegere, 2002; adapted from Fenson et al., 1993), as well as infants' cortical tracking of speech, but these will be reported in a separate paper.

2.1 | Participants

Thirty-three Dutch 10-month-old infants participated in the study (mean age = 308.82 days, age range = 291–326 days; 16 female). Fifty-seven additional infants were tested but were excluded from further analyses due to having too many flat or noisy channels (*n* = 6), not having at least 10 artefact-free trials per condition (*n* = 42), technical

TABLE 1 An example of an experimental block (English translations in parentheses, with the familiarised target word **bold and underlined**, and the novel control word underlined).

Familiarisation phase
1. Er zitten **cello's** in het orkest. *(There are cellos in the orchestra.)*
2. Goede **cello's** zijn van hout gemaakt. *(Good cellos are made of wood.)*
3. Ik hoorde vanochtend **cello's**. *(I heard cellos this morning.)*
4. Met de pauken spelen vaak de **cello's** mee. *(The cellos often play along with the timpani.)*
Test phase
1. **Cello's** *(cellos)*
2. Tuba's *(tubas)*

issues (*n* = 4) or refusal to wear the cap or excessive fussing (*n* = 5). The planned sample size was 48 infants (determined based on a power analysis performed using G*Power 3.1 (Faul et al., 2009) to obtain statistical power at the 0.8 level with *f* = 0.25). As specified in our pre-registration (see above), we aimed to recruit 90 infants to allow for 45%–50% attrition, resulting in 45–50 infants whose data could be used for the EEG analyses. (Note that due to a higher attrition rate, the included sample size of 48 infants was not achieved, and because of time limitations due to the effect of the COVID-19 pandemic on testing, we did not include additional participants after our original sample of 90 participants.) All infants were born full-term and were reported to have normal development, sight, and hearing. Infants came from monolingual Dutch families, who reported no neurological or language problems in the immediate family. Participants were recruited from the Nijmegen Baby and Child Research Center database. The study was approved by the Ethical Board of Social Sciences, Radboud University, Nijmegen. Parent(s) or caregiver(s) signed an informed consent form for the participation of their child in the study, in accordance with the Declaration of Helsinki, and were offered a choice between 20 Euros and a book as a token of appreciation.

2.2 | Materials

Materials consisted of blocks of audio-visual familiarisation sentences (four sentences per block) followed by isolated audio-only test words (see Table 1 for an example block, and Supplementary Materials D, Table S4 for the full set of stimulus materials).

2.2.1 | Familiarisation stimuli

To create the experimental materials, 90 low-frequency disyllabic trochaic Dutch words were selected, which had a frequency lower than 13 per million in the CELEX database (Baayen et al., 1995). These 90 words were combined into 30 triads, which were used as test words and to create the familiarisation sentences. For the familiarisation phase, combinations of sentences ('familiarisation blocks') were created, each comprising four sentences in which one of the target words was repeated. The same combinations of sentences were used for each item in each triad, resulting in 30 * three items = 90 blocks, consisting of 90 blocks * four sentences = 360 sentences. The sentences consisted of 8–12 syllables. In each block, the target word was the second or third word in three of the four sentences, and the last word in one sentence (in either the second or the third sentence). The familiarisation stimuli consisted of 90 videos of a female Dutch actor, speaking either with direct gaze or averted gaze. To ensure that the speech properties remain the same across conditions, videos were simultaneously recorded from three angles using three different cameras as the actor was reciting the 360 sentences: (1) speaker looking directly at the camera in the middle; (2) speaker's head averted at an approximately 20° angle to the left; and (3) speaker's head averted at an approximately 20° angle to the right. The actor was instructed to speak in a lively, infant-directed manner, and looked at a picture of an infant during stimulus recording. Stimuli were recorded using Adobe Audition, and were processed using Adobe Premier Pro for video editing, and Praat (Boersma & Weenink, 2021) for audio editing. The mean sentence duration was 3197 ms (*SD* = 507 ms), the inter-sentence interval was approximately 1500 ms, and the mean duration of each familiarisation block was 18.9 s. The mean target word duration in sentences was 728 ms (*SD* = 144 ms).

2.2.2 | Test stimuli

For the test phase, the 90 experimental words were recorded separately in isolation. The mean target word duration was 911 ms (*SD* = 127 ms). The single words were recorded in the same recording session as the familiarisation stimuli, using the three-camera setup, but only the audio stimuli were used in the experiment. The audio stimuli were processed using Praat (Boersma & Weenink, 2021).

2.3 | Design

Each participant was exposed to 60 experimental blocks. Each block consisted of a familiarisation phase (video; four sentences with one repeated target word) followed by the test phase (audio; two single words). In the test phase, one single word was the familiarised word the infant had just heard during the familiarisation phase, and the other word was a novel control word which infants had not been familiarised with (order counterbalanced; see Table 1 for an example block). The single words in the test phase were not accompanied by any visual cues to specifically test for the effects of processing in the familiarisation phase, that is indexed by the word segmentation effect in the test phase, without it being influenced by the online processing effects of the combination of speech and gaze.

In the second half of the experiment (trials 31–60), infants were shown the same familiarisation videos with the same target word as in the first half, but using a different control word in the test phase. The blocks were presented in a pseudo-randomised order, ensuring that there were at least 10 intervening blocks between the same blocks in the first and second repetition.

So that the first and second half of the experiment used different control words, three versions (A, B, C) were created out of the 90 familiarisation blocks containing the 90 words, resulting in 30 blocks in each version. The target word in one version served as the control word in the other two versions. For instance, an infant who was familiarised with Version A heard the control words from Version B in the test phase of the first half and from Version C in the second half.

Each participant was presented with 30 blocks in the direct gaze condition, and 30 blocks in the averted gaze condition. The speaker's gaze direction for the averted gaze condition (looking at left or right) was kept constant within participants but was counterbalanced between participants, to ensure that the visual features of one side of the speaker's face did not have any effects. The gaze condition (direct/averted) was altered every 2–3 trials. The order of the blocks as well as the familiarisation versions (A, B, C) were pseudo-randomised and counterbalanced between participants. Twelve lists were created to counterbalance the order of presentation, gaze condition (that is, one block that was presented with direct gaze in half of the lists was with averted gaze in the other half) and the different versions (A, B, C).

2.4 | Procedure

When the infant and parent came into the lab, the infant was invited to play on a play mat, while one experimenter briefed the parent about the study procedure. The infant's head circumference was measured, and the correct-sized EEG cap was pre-gelled. The EEG cap was then fitted on the infant's head, electrode impedances were checked, and additional gel was added if necessary.

After capping, the infant was seated in their parent's lap in an electrically shielded and sound-attenuated testing booth, approximately 70 cm away from a 24-inch display monitor. Audio stimuli were presented over two loudspeakers at approximately 65 dB. The videos were displayed at the centre of the screen (20 × 20 cm). Stimuli were presented using Presentation (Version 20.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com).

The experiment started with an attention-getter to direct infants' attention to the screen. Then, two silent baseline videos of the speaker looking at the infant, and looking away from the infant were presented, each for 10 s, with a 1000-ms interval in between (order counterbalanced). After that, the first experimental block began. In each block, one familiarisation video was first presented. After approximately 1500 ms, this familiarisation video was followed by the presentation of one target word and one control word, presented with an interval of 1500 ms between words (audio-only; order counterbalanced). The inter-trial interval between the offset of the test phase of one block and the onset of the familiarisation video of the next block was 3000 ms. Short attention getters were presented in a pseudo-randomized order, between every 4 to 5 blocks.

During the experiment, the infant was given silent toys to play with, and/or breadsticks, if they became restless. Parents listened to masking music through noise-cancelling closed-ear headphones. The experimenter ran the experiment and EEG acquisition from outside of the booth. The sessions were video-recorded using a CCTV video camera and a webcam for offline coding of infants' looking behaviour. Breaks were taken if the infant became fussy, during which the experiment was paused and infants could watch a silent cartoon. The session was stopped if the infant became distressed, or was no longer attending to the screen. The experiment lasted about 25–30 min, and the whole session lasted about an hour. During the experiment, the institution's COVID-19 measures were applied. The experimenters and parents wore face masks during the session, and the experimenters additionally wore face shields while fitting the cap.

2.5 | Looking times

To assess infants' looking times in the familiarisation phase, infants' looking behaviour during the presentation of each block was manually coded using ELAN (version 6.3, 2022). Infants' looks to the screen and looks away from the screen were coded frame-by-frame. We calculated infants' attention as indexed by the proportion of looking time to the video per block:

$$Attention = \frac{total\ looking\ duration}{duration\ of\ the\ familiarisation\ phase}\ \%$$

For the ERP analyses, we excluded trials in which infants attended to the screen for less than 25% of the duration of the familiarisation phase, similar to previous eye-tracking and EEG studies that suggested similar exclusion criteria (e.g. 15% in LoBue et al., 2017; 20% in Taylor & Herbert, 2013; also see Tan et al., 2022). This means that if the infant attended to the screen for less than 25% of the video duration time in the familiarisation phase, the subsequent ERP test trials that were part of the same block were discarded. While the looking time measure was primarily used as an exclusion criterion, we also assessed differences in infants' visual attention to the stimuli in the Direct and Averted gaze conditions.

2.6 | EEG recordings and processing

EEG was recorded from 32 Ag/AgCl electrodes (ActiCAP) placed according to the International 10–20 system, using BrainAmp DC and Brain Vision Recorder software (Brain Products GmbH, Germany). FCz was used as the online reference. EEG was recorded from the following electrodes: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10, P7, P3, Pz, P4, P8. EOG was recorded from the electrode above (Fp1) and an additional electrode placed below the eye (VEOG), and additionally from the two electrodes at the outer canthi of the eyes (FT9, FT10). Besides the two mastoid electrodes in the cap (TP9, TP10), two additional loose electrodes were placed directly on the mastoid bones ('TP9L', 'TP10L'). EEG data were recorded with a sampling rate of 500 Hz, using an online time cut-off of 10 s and a high cut-off of 1000 Hz. Impedances were typically kept under 25 Ω.

EEG data were processed and analysed using Fieldtrip toolbox for EEG/MEG-analysis (Oostenveld et al., 2011) in MATLAB (version

2020b). Eye and noise components in the data were identified using Independent Component Analysis (ICA; Makeig et al., 1996). Prior to ICA, data were filtered with a Hamming windowed Butterworth high-pass filter of 0.1 Hz (−12 dB/oct) and a low-pass filter of 30 Hz, and cut into 1-s segments. Data were visually inspected and channels and data segments with flat channels or large artefacts (exceeding 150 μV for EEG channels, 250 μV for EOG channels) were excluded. Next, Independent Component Analysis with Infomax ICA (Bell & Sejnowski, 1995) was applied. Eye movement and noise components (i.e. components with activity on a single electrode, heart beat components) were identified by visual inspection of the components and data. On average, 2.87 eye (range: 1–5) and 4.09 noise components (range: 2–8) were removed. Then, for the test phase, raw data were epoched from 200 ms before to 900 ms after the critical word onset (i.e., the novel and familiar words in the test phase). For the familiarisation phase, time-locked data were created by epoching the raw data from 200 ms before to 800 ms after the onset of the target words within the sentences, given the shorter mean target word duration within sentences. Data were filtered again from 0.1 to 30 Hz using the abovementioned low- and high-pass filters, and the identified eye movement and noise components were removed from the data. The cleaned EEG data were re-referenced to the linked mastoids (alternatively, if the mastoids were not usable, TP9 and TP10 or a bilateral combination thereof was used). Data were baseline-corrected using the 200 ms prior to the critical word onset as the baseline. Then, trials with amplitudes exceeding ±150 μV were automatically rejected. Additionally, trials in which infants' attention during the familiarisation phase (as defined by the proportion of looking time to each trial) was less than 25% of the trial duration were discarded. Six datasets were removed because they had more than four noisy or flat channels. Forty-two datasets were removed because they had fewer than 10 trials per condition in the test phase after the artefact and looking time exclusion (note that the exclusion rate is higher than what is usually reported in the literature because our exclusion criteria required infants to have at least 10 artefact-free trials out of a maximum of 30 per condition for all four conditions). Of those 42 infants, four were excluded because they did not have enough trials in each condition after further exclusion based on looking behaviour (i.e. they would have been included if we did not exclude trials due to looking behaviour). Bad channels in the remaining datasets were repaired with a spherical spine interpolation (Perrin et al., 1989), which is implemented in the FieldTrip toolbox. Single-trial mean EEG amplitudes of pre-defined time-windows (see below) were extracted for all included trials for statistical analyses. Participant ERPs were made by averaging over the relevant trials per condition. For illustration purposes, grand average ERPs were also created.

2.7 | ERP analyses

2.7.1 | Test phase

For the test phase, we computed the mean amplitude of the two pre-defined time windows per trial for each infant per each familiarity and gaze condition, resulting in four conditions: familiar word following the direct gaze condition in the familiarisation phase (henceforth Direct Familiar), novel word following the direct gaze condition (Direct Novel), familiar word following the averted gaze condition (Averted Familiar) and novel word following the averted gaze condition (Averted Novel). As mentioned before, each infant had a minimum of 10 artefact free trials per condition out of a maximum of 30 trials per condition (Direct Familiar: $M = 15.6$, $SD = 3.9$; Direct Novel: $M = 15.6$, $SD = 3.5$, Averted Familiar: $M = 14.7$, $SD = 4.1$, Averted Novel $M = 15.9$, $SD = 4.4$). A 2 (familiarity) × 2 (gaze) repeated-measures ANOVA showed a significant effect of familiarity on the mean number of included trials ($F(1,32) = 6.62$, $p = 0.015$), with more trials in the novel than the familiar condition, but note that the difference was very small (15.2 vs. 15.8). The main effect of gaze and the interaction between gaze and familiarity was not significant.

The time windows and region of interest were selected a priori based on existing literature on infants' word familiarity effect (see pre-registration: https://aspredicted.org/blind.php?x=VSG_ZNB). Accordingly, two time windows of interest (250–500 ms; 600–800 ms) were defined (Kidd et al., 2018; Snijders et al., 2020). Our main region of interest was the left-frontal electrodes (F7, FC5, F3), given that the ERP word familiarity effect is most consistently observed and reported over the left-frontal regions (Junge et al., 2012, 2014; Kidd et al., 2018; Kooijman et al., 2005, 2009, 2013; Männel & Friederici, 2013; Snijders et al., 2020). As we were assessing the ERP word familiarity effect, our main effect of interest is Familiarity and the interaction effects of our other variables (Time, Gaze) with Familiarity, rather than the main effects of Time and Gaze.

To this end, we analysed the trial-level test phase data at our a priori selected time windows (250–500 and 600–800 ms) and region of interest (left-frontal electrodes) using a linear mixed-effects model. The model included time, familiarity and gaze direction as fixed effects, and by-item and by-participant random effects. The categorical variables Time (early: 250–500 ms/late: 600–800 ms), Familiarity (familiar/novel) and Gaze (direct/averted) were contrast coded with the contrasts (−0.5, 0.5), with early, novel and direct as the reference levels coded as −0.5. We started with the maximal random effects structure (Barr et al., 2013) including both random intercepts and random by-item and by-participant slopes for our predictor variables. As the maximal model did not converge, we reduced our model following a stepwise removal procedure of the random effects, and fitted the most parsimonious model which included by-item and by-participants random intercepts. The final model contained mean amplitude as the outcome variable, time, familiarity and gaze as the predictors, and participant and item as random intercepts (meanAmplitude ∼ time * familiarity * gaze + (1|participant) + (1|item)) and was fitted with restricted maximum lixelihood (REML). Statistical analyses were performed using the lme4 (Bates et al., 2015) and lmerTest packages (Kuznetsova et al., 2017) in R (version 4.2.2; R Core Team, 2022).

Furthermore, we explored possible effects of our main experimental variable, the direction of the speaker's gaze, outside of the pre-defined regions of interest. For this, we used non-parametric cluster-based randomisation tests (Maris & Oostenveld, 2007) to assess the differences in the ERP familiarity effect between the Direct and Averted gaze conditions. Cluster-based permutation analysis makes use of dependent samples *t*-tests to identify electrodes that exceed a threshold alpha level (0.05), and forms clusters by grouping together neighbouring electrodes and time points that exceed this threshold. Then, a cluster-level statistic is calculated by summing the t-statistics of all identified clusters. Using Monte Carlo re-sampling (with 1000 permutations), a reference randomisation distribution for surrogate data is created by randomly re-assigning data from different conditions, which is then compared to the observed cluster-statistics to obtain a Monte Carlo *p*-value. If the *p*-value of this test statistic is below a critical threshold (e.g. 0.05), it can be concluded that the data in the two conditions are significantly different. Due to the fact that all points of interest (e.g. electrodes) can be tested with one statistical test, cluster-based randomisation analysis controls for the multiple-comparisons problem that stems from the multi-dimensionality of the EEG data (Maris & Oostenveld, 2007).

To investigate the effects of the speaker's gaze direction on infants' word segmentation, we conducted two exploratory cluster-based permutation tests comparing the difference waves of the word familiarity effect (Familiar minus Novel) in the two conditions (Direct vs. Averted) by assessing all EEG electrodes (F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8), averaging over time for one time window in each test (250–500 and 600–800 ms).

2.7.2 | Familiarisation phase

We also conducted an exploratory (i.e. not pre-registered) analysis of the word familiarity effect in the familiarisation phase, by comparing the same infants' ERPs for the first two occurrences of the target words within the sentences to the last two occurrences. Infants had an average of 33.8 trials (*SD* = 8.0) for the first two repetitions in the Direct gaze condition, 33 trials (*SD* = 8.3) for the first two repetitions in the Averted gaze condition, 31.7 trials (*SD* = 7.9) for the last two trials in the Direct gaze condition, and 31.8 trials (*SD* = 7.4) for the last two trials in the Averted gaze condition out of a maximum of 60 trials per condition. A 2 (repetition) × 2 (gaze) repeated-measures ANOVA showed a significant effect of repetition on the mean number of included trials ($F(1,32) = 10.37$, $p = 0.003$), with fewer trials at the end of the familiarisation phase than at the beginning. The main effect of gaze and the interaction between gaze and repetition was not significant.

We computed the mean amplitude of the ERP response to each occurrence of the target word within sentences, in the 250–500 and 600–800-ms time windows for each gaze condition. First, we tested the word familiarity effect over the left-frontal region of interest using a linear mixed-effects model. Again, our main effect of interest was the effect of Familiarity (here: Repetition, comparing the 1st and 2nd tokens to the 3rd and 4th), and the interaction of Repetition with Time and Gaze. The predictors were contrast coded with the contrasts (−0.5, 0.5), with 1st and 2nd tokens, early, and direct as the reference levels (coded −0.5). In line with our planned analysis of the test phase, we started with the maximal random effects structure, and fitted the most parsimonious model after stepwise removal of the random effects, resulting in random slopes for both participant and item, with mean amplitude as the outcome variable, Repetition (1st and 2nd/3rd and 4th), Time (early: 250–500 ms/late: 600–800 ms) and Gaze (direct/averted) as the predictors, and by-participant and by-item random effects.

In addition, we further explored the effects of gaze over all electrodes by comparing infants' word familiarity effect (difference between 1st and 2nd tokens and 3rd and 4th) in the two conditions (Direct vs. Averted) using cluster-based permutation tests, averaging over time for each time window (250–500 and 600–800 ms) in each test.

3 | RESULTS

3.1 | Looking times

We first tested whether there were differences in attention to the familiarisation stimuli across the direct and averted gaze conditions in the included trials. We compared infants' mean looking times per trial during the familiarisation phase in the Direct and Averted gaze conditions. A paired-samples *t*-test indicated that infants' looking times did not differ significantly between the Direct (*M* = 13.58 s [attention in percentage: 71.70%], *SD* = 4.61 s [24.13%]) and Averted gaze (*M* = 13.57 s [71.59%], *SD* = 4.74 s [24.67%]) conditions, $t(32) = -0.17$, $p = 0.9$. Figure 1 illustrates looking times per trial in the two conditions.

3.2 | ERP familiarity effect in the test phase and the effects of gaze

The ERPs over the left-frontal electrodes, and the topographical distributions of the overall difference between the familiar and novel words in the two time windows, averaged over the gaze conditions, are shown in Figure 2. Our pre-registered analysis focused on the left-frontal electrodes, but a visualization of the ERPs for the four conditions across all electrodes can be found in Supplementary Materials (Section C; Figure S3).

To investigate the role of the speaker's gaze direction on infants' word segmentation performance, we used a linear mixed-effects model to evaluate whether the mean amplitude of infants' responses to the familiar and novel words in the test phase was modulated by the speaker's gaze direction in the familiarisation phase. Table 2 shows the mean (SD) amplitudes of ERP to familiar and unfamiliar words, and Table 3 presents an overview of the full model statistics. As our main research question deals with the word familiarity effect and the effects of the speaker's gaze on word familiarity, here we focus only on reporting the main effects and interactions relevant to our predictions. The results suggested a significant effect of Familiarity, as predicted, with

TABLE 2 Test phase: means and standard deviations of the mean amplitudes of ERPs to the familiar and novel words (split per gaze condition) in the two time windows and the average of both time windows.

	Mean amplitude (μV)					
	250–500 ms		600–800 ms		Average	
	Mean	SD	Mean	SD	Mean	SD
Familiar	4.83	24.38	2.35	28.53	3.59	26.56
Novel	5.55	24.89	5.17	28.85	5.36	26.94
Direct Familiar	5.38	24.34	2.97	28.30	4.17	26.41
Direct Novel	5.45	25.07	4.71	29.10	5.08	27.15
Averted Familiar	4.25	24.43	1.68	28.79	2.96	26.71
Averted Novel	5.65	24.74	5.62	28.62	5.64	26.74

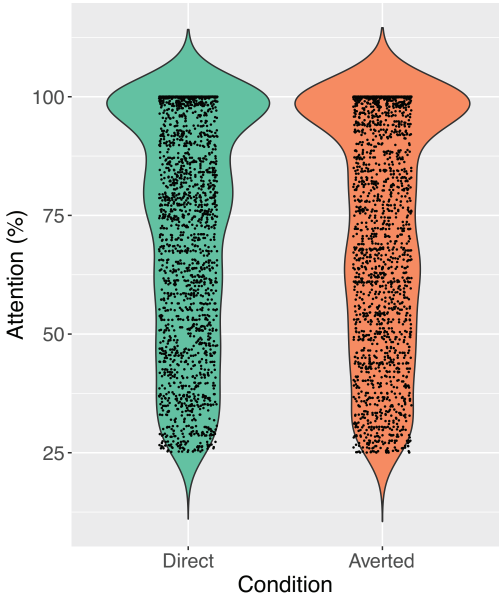

FIGURE 1 Looking times (percent of looking to the screen in each trial) during the familiarisation phase of the included ERP test trials in the Direct and Averted gaze conditions. The individual data points show each infants' proportion of looking per trial. ERP, event-related potential.

TABLE 3 Familiarity effect in the test phase: Model output for final model (reference levels: Early [250–500 ms] for Time, novel for Familiarity, direct for Gaze).

	Mean amplitude			
Predictors	*Estimate*	*SE*	*CI*	*p*
(Intercept)	4.39	0.98	2.47–6.30	<**0.001*****
Time	−1.43	0.82	−3.04–0.18	0.081
Familiarity	−1.92	0.82	−3.54– −0.31	**0.020***
Gaze	−0.58	0.83	−2.20–1.04	0.484
Time * Familiarity	−2.10	1.64	−5.32–1.11	0.200
Time * Gaze	0.27	1.64	−2.95–3.49	0.869
Familiarity * Gaze	−2.12	1.65	−5.35–1.12	0.200
Time * Familiarity * Gaze	−0.87	3.28	−7.31–5.57	0.791
Random effects				
σ^2	687.09			
$\tau_{00\ \text{item}}$	7.51			
$\tau_{00\ \text{participant}}$	22.83			
ICC	0.04			
$N_{\text{participant}}$	33			
N_{item}	90			
Observations	4082			
Marginal R^2/ Conditional R^2	0.003/0.045			

$^{***}p < 0.001$, $^{*}p < 0.05$.

a decrease in mean amplitude to Familiar compared to Novel trials ($ß = -1.92$, $SE = 0.82$, $t = -2.33$ $p = 0.020$). Visual inspection of the grand average waveforms for the Familiar and Novel conditions suggests that this familiarity effect becomes larger with time, thus appears to be larger in the 600–800-ms time window (Figure 2a). However, the interaction of Time and Familiarity was not significant ($ß = -2.10$, $SE = 1.64$, $t = -1.28$, $p = 0.200$).

The Familiarity by Gaze interaction was not significant ($ß = -2.12$, $SE = 1.65$, $t = -1.28$, $p = 0.200$), nor was the main effect of Gaze ($ß = -0.58$, $SE = 0.83$, $t = -0.70$, $p = 0.484$) or either of the other effects, including the interaction between Time, Familiarity and Gaze ($ß = -0.87$, $SE = 3.28$, $t = -0.26$, $p = 0.791$). Figure 3 shows the grand average waveforms over left frontal electrodes for the Familiar and Novel critical words following the Direct and Averted Gaze familiarisation conditions, and Figure 4 illustrates the topographical distribution of the familiarity effect in the Direct and Averted conditions in the two time windows.

We also tested whether infants' variability in attention (indexed by the proportion of looking times to the screen per trial) explained variability in their ERP word familiarity effect by adding Attention per trial to the model as a fixed factor. The results of this model indicated that Attention did not significantly predict the mean amplitude change ($ß = 0.04$, $SE = 0.02$, $t = 1.91$, $p = 0.057$), while Familiarity still had a significant effect, again with a decrease in mean amplitude to Familiar

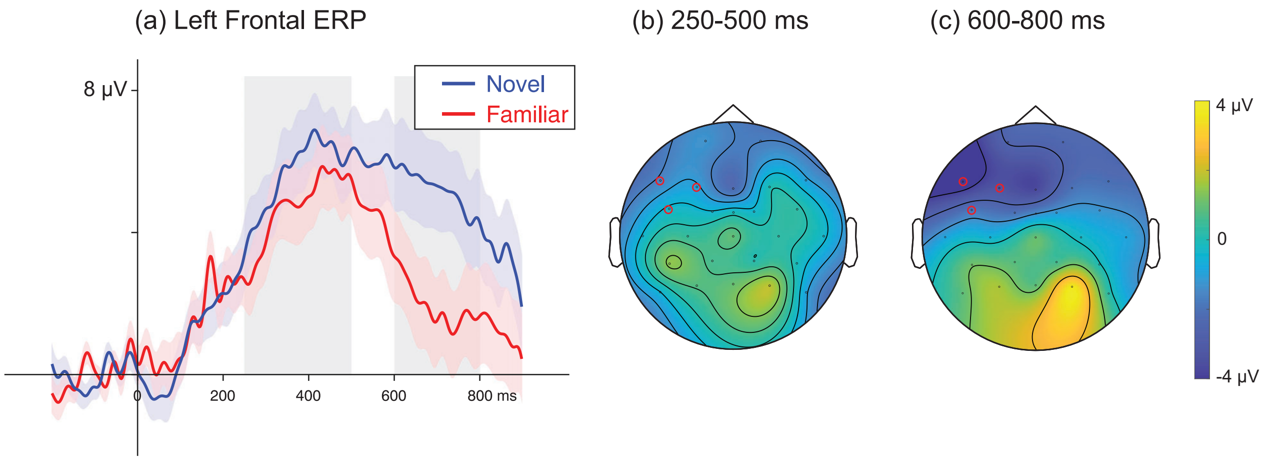

FIGURE 2 (a) Event-related potentials over the left-frontal electrodes (F7, F3, FC5) in the test phase, for the familiar and novel words, averaged over the gaze conditions. The grey areas indicate the time windows of interest (250–500, 600–800 ms) and the shaded areas around the waveforms represent the standard error of the mean. Right panel: Topographic isovoltage maps of the familiarity effect, in the (b) 250–500 ms and (c) 600–800-ms windows, averaged over the Direct and Averted gaze conditions. The red circles indicate the left-frontal region of interest electrodes.

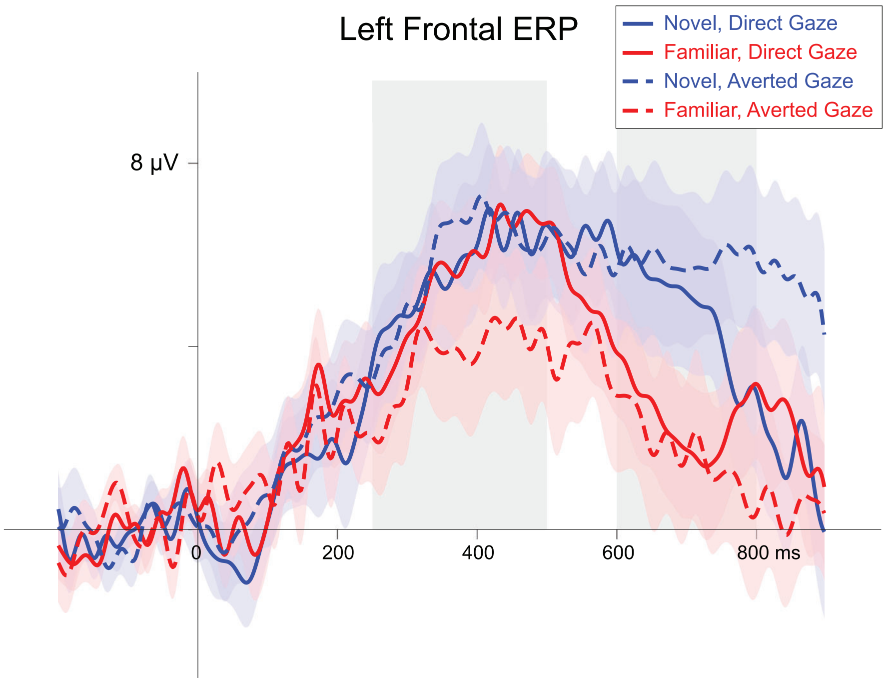

FIGURE 3 Event-related potentials over the left-frontal region of interest electrodes (F7, F3, FC5) in the test phase, for the Direct Novel, Direct Familiar, Averted Novel and Averted Familiar conditions. The grey areas indicate the pre-defined time windows of interest (250–500, 600–800 ms) and the shaded areas around the waveforms represent the standard error of the mean.

compared to Novel trials ($\beta = -1.93$, $SE = 0.82$, $t = -2.34$, $p = 0.019$). None of the other main effects and interactions were significant. Moreover, adding Attention as a fixed factor did not significantly improve model fit compared to the previously fitted model ($\chi^2 = 3.65, p = 0.056$; AIC fitted model = 38,353, AIC model with Attention = 38,351). Therefore, we did not include Attention per trial as a predictor in our main model (see Supplementary Materials A, Table S1 for the full model output).

3.3 | Exploratory analyses

3.3.1 | Cluster-based permutation analysis of the gaze effect in the test phase

In our planned analyses, we selected the time windows and regions of interest for the ERP familiarity effect a priori based on previous literature. However, our design differed from previous ERP word

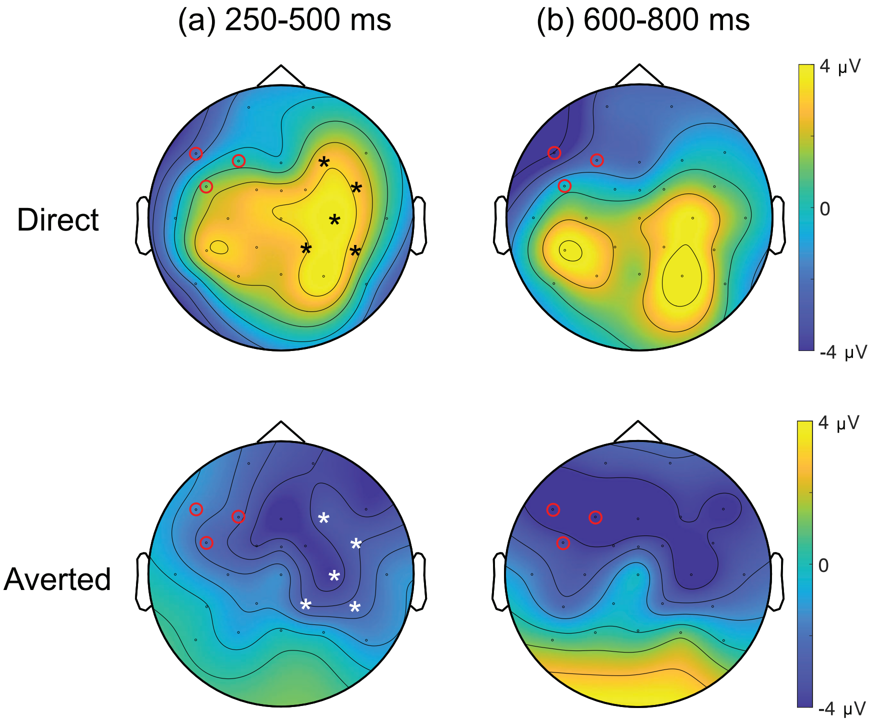

FIGURE 4 Topographic isovoltage maps of the familiarity effect, in the (a) 250–500 ms and (b) 600–800-ms windows, in the Direct (top) and Averted (bottom) gaze conditions in the test phase. The left-frontal electrodes are marked with red circles. The electrodes that are included in the significant cluster of the cluster-based permutation test testing the difference between the word familiarity effect in the Direct and Averted gaze conditions are highlighted with asterisks in the 250–500-ms window (cluster $p = 0.048$, cluster electrodes: (F4, FC6, C4, CP2, CP6)).

segmentation designs in that we used audio-visual stimuli and assessed the effect of the speaker's gaze. Thus, it might be the case that there were differences in the ERP familiarity effect that fell outside the traditional region of interest. To investigate this, we explored the effect of gaze on different regions by means of cluster-based permutation analyses, assessing all EEG electrodes at once, comparing the word familiarity effect (calculated by subtracting the responses to the novel words from the responses to the familiar words for each gaze condition) for the Direct and Averted conditions in the two time windows. We assessed the 250–500 and 600–800 ms time windows in two separate tests, averaging over time.

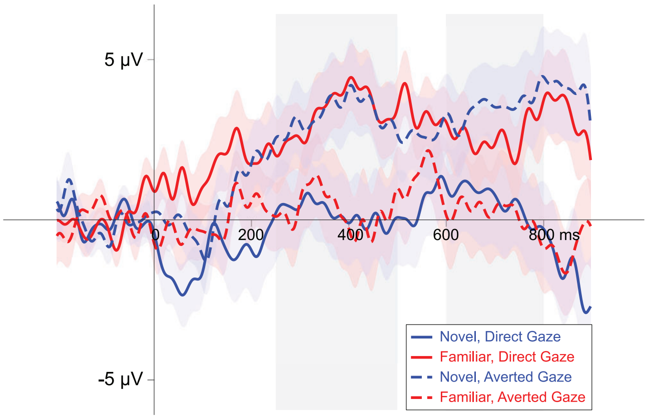

FIGURE 5 Event-related potentials over the cluster electrodes identified in the 250–500-ms time window in cluster-based permutation analyses (F4, FC6, C4, CP2, CP6). The waveform depicts the ERPs in the test phase for the Direct Novel, Direct Familiar, Averted Novel and Averted Familiar conditions, and the shaded areas around waveform represent the standard error of the mean. The grey areas indicate the pre-defined time windows of interest (250–500, 600–800 ms).

The topographic isovoltage maps of the ERP word familiarity effect in the Direct and Averted Gaze conditions (Figure 4) suggest that, besides the negative familiarity effect over left-frontal electrodes, there also might be a positive familiarity effect over more posterior electrodes in the Direct gaze condition, while the negative familiarity effect might be more extended towards frontocentral electrodes in the Averted Gaze condition. Indeed, the cluster-based permutation tests revealed significant differences in the Familiarity effect between the Direct and Averted Gaze conditions. One positive cluster was identified in the 250–500-ms window (cluster $p_{corrected} = 0.048$). No clusters were identified in the 600–800-ms time window. The difference was most prominent over the right frontal and central electrodes (see Figure 4). To follow up on the difference between the conditions in the early time window, we extracted the mean ERP between 250 and 500 ms, using the identified cluster electrodes (F4, FC6, C4, CP2, CP6). Post-hoc analyses using paired t-tests suggested a significant positive word familiarity effect in the Direct Gaze condition, with Familiar words ($M = 3.32$, $SD = 5.44$) showing a larger amplitude compared to Novel words ($M = 0.2$, $SD = 6.56$), $t(32) = 2.23$, $p = 0.033$. For Averted Gaze, the negative word familiarity effect was not significant over these right frontal and central electrodes in the 250–500-ms time window ($M_{familiar} = 0.27$, $SD_{familiar} = 6.95$; $M_{novel} = 3.29$, $SD_{novel} = 6.30$; though note $t(32) = -1.81$, $p = 0.08$). Figure 5 illustrates the average waveforms over the cluster electrodes in the 250–500-ms time window.

3.3.2 | ERP familiarity effect in the familiarisation phase and the effects of gaze

We conducted a follow-up analysis of the ERP responses to the target words during the familiarisation phase to better understand the word familiarity effect. We compared infants' left-frontal ERP responses to the last two occurrences of the target word (familiarised) with those to the first two occurrences (unfamiliarised) by means of a linear mixed model. The results indicated that the mean amplitude of infants' ERP response did not significantly change between the first and last two repetitions of the target word, meaning that infants did not show the negative-going word familiarity effect in the familiarisation phase ($ß = 0.94$, $SE = 0.55$, $t = 1.72$, $p = 0.085$). The main effect of Gaze was not significant ($ß = 0.12$, $SE = 0.55$, $t = 0.21$, $p = 0.833$), even though the first two repetitions in the Direct Gaze condition seem to have elicited a more negative response. A significant main effect of Time ($ß = -1.14$, $SE = 0.54$, $t = -2.09$, $p = 0.037$) showed that infants' mean amplitudes decreased from the early to the late time window. The interactions of Repetition and the other predictors were not significant (see Supplementary Materials B; Tables S2 and S3 for the descriptive statistics and the full model output, Figure S1 for the ERPs and topographical distribution of the word familiarity effect averaged over the gaze conditions, and Figure S2 for the ERPs for the two gaze conditions).

The cluster-based permutation comparing the word familiarity response (computed as the difference wave between the last two and first two repetitions of the target word in the familiarisation phase) in the direct and averted gaze conditions over the two time windows did not identify any clusters, meaning that no significant difference in word familiarity with gaze was observed at any electrode site during the familiarisation phase.

4 | DISCUSSION

It has been well-established from both behavioural and EEG studies that, within the first year of life, infants gradually master the ability to extract and recognise words within continuous speech, which is a fundamental skill for vocabulary development. So far, most studies on word segmentation have only focused on language as unimodal speech input. The natural mode of communication in many infant–caregiver interactions, however, is face-to-face interaction, in which ostensive signals, such as mutual gaze, are used abundantly by caregivers. Such cues may increase infants' attention to the speaker and to the input provided by the speaker, thereby facilitating speech processing tasks, such as segmentation from continuous speech.

This study investigated whether ostensive communication, as signalled by the speaker's use of eye contact, might facilitate infants' word segmentation. We analysed infants' ERP responses in an EEG familiarisation-then-test paradigm. We predicted that infants would show the word familiarity effect and that this effect would be enhanced (i.e. the familiar-novel word difference would be larger), when speech with which they were familiarised was ostensive, as signalled through the speaker's use of direct eye gaze. Our results indicated that infants successfully segmented words from continuous speech in the audio-visual stories, differentiating between familiarised novel words in the test (but not the familiarisation) phase. We did not find the predicted effect of gaze on infants' word familiarity responses over our left-frontal region of interest. However, our exploratory analyses revealed a central-posterior and right fronto-central positive familiarity effect that differed between the direct and averted gaze conditions.

In line with previous research, the 10-month-olds in our study had a negative-going ERP response to familiarised words compared to unfamiliarised (novel) words over left-frontal electrodes (Goyet et al., 2010; Junge et al., 2012; Kidd et al., 2018; Kooijman et al., 2005; Männel & Friederici, 2013) in the test phase. This negative-going response is usually interpreted as a mature response and has been linked to later language skills, with studies reporting significant correlations between the magnitude of the word familiarity effect and later receptive and expressive vocabulary (Junge et al., 2012; Kooijman et al., 2013; Von Holzen et al., 2018). The main effect of familiarity that we observed did not differ between the two time windows, although the effect appeared to be more pronounced in the late (600–800 ms) time window. While earlier ERP responses are commonly reported in paradigms in which infants are familiarised with isolated single word tokens and then tested on their ability to recognise this word in continuous speech (e.g. Kooijman et al., 2005, 2013), paradigms similar to our design, which present the target word in continuous speech, requiring infants to segment first and then recognise this token in the test phase (Junge et al., 2012), also have elicited later responses. In fact, Junge et al. (2012) reported that 10-month-old infants who had a larger vocabulary at 12 months succeeded in segmenting target words from continuous speech and showed a long-lasting word familiarity effect, which increased over time, while infants with lower later vocabulary scores only displayed a short-lived early familiarity effect, and only in the condition in which they were familiarised with single words.

Contrary to our predictions, our planned analyses did not show a significant difference between the familiarity effect in the direct and averted gaze conditions over left-frontal electrodes. We predicted, based on Natural Pedagogy theory (Csibra & Gergely, 2009), that the ostensive communication mode, signalled through the speaker's use of direct eye gaze during the familiarisation phase, would facilitate infants' word segmentation in the test phase, resulting in an enhanced negativity. There was no evidence for such a facilitatory role of eye contact on word segmentation over left-frontal electrodes.

However, while there was no effect of gaze on the word-familiarity effect over the left-frontal region of interest, the topography of the word familiarity effect in the direct versus averted gaze conditions pointed to an additional positivity in the direct gaze condition over centro-parietal and right fronto-central electrodes, besides the left-frontal negative effect. To investigate this finding further, we conducted exploratory follow-up analyses of the difference waves of the word familiarity effect in the direct and averted gaze conditions, assessing all electrodes at once in cluster-based permutation tests. The follow-up analyses suggested a significant difference between the two gaze conditions in the 250–500-ms time window, mainly over right fronto-central and centro-parietal electrodes. This difference was driven by a positive word familiarity effect in the early window in the direct gaze condition. In the 600–800-ms window, no differences between gaze conditions were identified. The early central and right fronto-central positive response in the direct gaze condition, partly overlapping with the left-frontal negativity, might have attenuated the left-frontal negativity to a degree, resulting in a smaller negative early ERP familiarity response between 250 and 500 ms over left-frontal electrodes.

We also conducted an exploratory follow-up analysis of the word familiarity effect in the familiarisation phase, during which infants were presented with four sentences containing the target word. Comparing the first occurrences of the target word to the last two, we found no evidence for the word familiarity effect in the familiarisation phase over the left-frontal region of interest. The direct and averted gaze conditions did not result in significantly different word familiarity effects in the familiarisation phase, neither over the left-frontal region of interest nor in the cluster-based permutation analysis testing all electrodes. This suggests that, even though infants showed the negative-going word recognition response in the test phase, they did not show a clear indication of word segmentation from continuous speech during familiarisation yet at the third and fourth occurrence of the word. This might be due to the difficulty of segmenting words that have not been presented in isolation before, as a familiarity response to words within utterances requires both the fast segmentation of the token from continuous speech and further recognising this token as a familiar word form (Junge et al., 2012; Kooijman et al., 2013). Moreover, both Junge et al. (2014) and Männel and Friederici (2013) show the word familiarity effect in infants only for later occurrences in continuous speech, that is, when comparing the first two repetitions to the seventh and eighth, but not the third and fourth. This suggests that infants show the word familiarity effect in continuous speech only *after* four occurrences, which might explain why we did not find this effect in the familiarisation phase within only four repetitions in the current study.

Overall, our results suggest that 10-month-old infants are able to recognise isolated word tokens that they segmented from continuous speech, after four repetitions in continuous speech, and that they can do this both with and without eye contact. In other words, eye gaze does not have a facilitatory effect, at least in the predicted left-frontal region. There are several possible explanations for this. First, this may be due to the 10-month-old infants in our sample already having mastered word segmentation and performing at ceiling, thereby not requiring any additional cues to help them segment words from continuous speech. This, however, is unlikely, as infants' mean amplitude differences in the 200–900-ms window ranged from −11.21 to 8.86 following a normal distribution, indicating that 10-month-olds in our sample did not uniformly show the negative-going word familiarity effect that is considered more mature, thus pointing against a ceiling effect. This variance is in line with previous studies, suggesting that infants can be classified into positive- and negative-responders based on the polarity of their ERP responses within the same age range (Kidd et al., 2018; Kooijman et al., 2013).

Another explanation might be that this difference was not observed over the left-frontal electrodes as we predicted, but was present in another region. This might be due to the audio-visual presentation mode that we used in the familiarisation phase, which differed from previous word segmentation studies that used an audio-only paradigm. As a result of this audio-visual (and social) familiarisation phase, we might expect to see differences in word familiarity that extend beyond the left-frontal regions. Indeed, while we did not observe any significant differences in infants' word familiarity effect over the left-frontal regions depending on the presence and absence of gaze cues, our exploratory cluster-based permutation analyses indicated that infants' word familiarity ERPs in the right fronto-central and centro-parietal regions were more positive-going when the speaker addressed them with eye contact, that is, after being familiarised with direct gaze compared to averted gaze, in addition to the left-frontal negativity. Interestingly, the broad distribution of the positivity resembles the topography of the word repetition effect observed in adults, which revealed a positive repetition effect over centro-posterior electrodes (Snijders et al., 2007). The difference in infants' ERP responses could occur because the use of eye gaze cued a more receptive state of information processing, as suggested by the Natural Pedagogy theory (Csibra & Gergely, 2009). It is, therefore, possible that ostensive communication, such as a speaker addressing infants with direct eye gaze, brings about differential processing of speech, reflected in the positive-going response in central and fronto-parietal regions for words familiarised during ostensive communication, similar to the topography of the emerging theta network in the first year of life (van der Velde et al., 2021).

It is also interesting that the presence of direct gaze led to a positive-going, not negative-going, response in the right fronto-central and centro-parietal regions. While the polarity of the ERP response does not necessarily reflect a novelty or familiarity preference, it is possible that the negative-going effect that is considered to be a more mature response indicates a preference for, or attention to, the encoding of novel words. The additional positivity over right frontal and central electrodes observed in response to the familiar test words in the direct gaze condition might be an indicator of infants' switching from a novelty-oriented to a familiarity-oriented processing mode, possibly because of the eye gaze of the speaker, leading infants to allocate more selective attention to the familiar items. However, infants' familiarity and novelty preferences, and whether and how they switch from a familiarity to a novelty preference have been a much-debated topic, and it is not entirely clear which factors drive a familiarity versus a novelty bias, making it difficult to draw robust conclusions. Overall, the results of all these exploratory analyses must be interpreted with caution, especially given that the study may be under-powered ($N = 33$ compared to planned 48; see Method section). Further work is necessary to explore these alternatives.

Finally, it is important to acknowledge the other sources of information that infants process along with the gaze cues when they see a talking face, such as the visual speech cues from the movement of the lips, jaw and larynx, which might enhance speech perception by providing predictive cues for the rhythmic structure of language (Fort et al., 2013; Moradi et al., 2013). These cues might be even more crucial for young infants who are starting to acquire the phonetic categories of their native language. Indeed, previous research has reported that infants and children show an audio-visual speech benefit, meaning that visual speech cues augment their speech perception (Lalonde & Werner, 2021). Five-month-old infants segmented words from the continuous speech of a target speaker in a multi-talker environment when they could see the target speaker's articulatory movements, but not when they were presented with a still picture of the speaker (Hollich et al., 2005). Another study found that 5-month-old infants' and adults' cortical tracking of speech was enhanced when speech was audio-visual compared to audio-only, whereas 4-year-old children did not show this audio-visual speech benefit (Tan et al., 2022). Furthermore, studies suggested a developmental shift in infants' looking behaviour at a talking face, with infants gradually shifting their attention to the speaker's mouth rather than to the eyes between 4 and 8 months, and returning to attending preferentially to the speaker's eyes by 12 months (Lewkowicz & Hansen-Tift, 2012; Tenenbaum et al., 2015). Therefore, it is possible that the 10-month-olds in our study did not benefit from the eye contact effect, as they attended preferentially to the mouth region of the speaker. This might also explain the observed difference (in the exploratory analysis) in the word familiarity effect in the direct and averted gaze conditions. In the direct gaze condition, the social information provided by the eyes competes with the temporal information provided by the mouth and the lips, which poses a challenge for word segmentation and thus results in a less mature ERP response, which is especially difficult for those infants who have not mastered the word segmentation skill yet. In the averted gaze condition, on the other hand, the lack of social gaze allows infants to focus primarily on the mouth region of the speaker, thereby maximising the chances of extracting the word tokens from continuous speech.

Adult studies have suggested that visual speech cues modulate auditory speech processing (Brown & Strand, 2019; Mitchel & Weiss, 2014; Saint-Amour et al., 2007; van Wassenhove et al., 2005). The word repetition effect in adults was modulated by speech modality in early stages of lexical processing (Basirat et al., 2018). Critically, this modulation appears to be linked to listeners' attention to the articulatory movements of the speaker. In a word segmentation task, adult listeners tend to spend more time looking at the speaker's mouth relative to the eyes while being familiarised with the artificial language, and their performance is, thus, significantly better in the audio-visual condition compared to the audio-only condition (Lusk & Mitchel, 2016). Moreover, listeners' shifts in gaze duration on the mouth region of the speaker have been associated with better performance on the word segmentation task (Lusk & Mitchel, 2016). However, the precise mechanisms of the eye contact effect and the auditory-visual speech benefit, especially in infants, remain to be explored. Further research should assess the role of visual speech cues on infants' word segmentation abilities using naturalistic paradigms. Accordingly, the eye contact effect should be investigated in different age groups, as infants might differ in the extent to which they benefit from the eye contact effect depending on which cues they need for speech segmentation. Ten-month-old infants might rely primarily on the visual speech cues, whereas younger or older infants might attend more to the gaze cues provided by the speaker. Future work looking at the effects of eye gaze and different types of visual speech cues separately is needed to tease apart the effects of gaze and visual speech cues on infants' speech processing.

In conclusion, this study provided evidence that 10-month-old infants can segment words from audio-visual speech and recognise those words in isolation, irrespective of the ostensiveness of the speech register, as seen in their ERP responses. While our results did not suggest a role of speaker's gaze on infants' word segmentation abilities in the left-frontal region of interest, we observed differential processing of familiar and novel words as a function of speaker's gaze mainly over central and right fronto-central regions. Future research is needed to better understand possible explanations of this effect.

ACKNOWLEDGEMENTS

We would like to thank all infants and their parents for their participation in this study, and the research assistants for their help with the creation of stimulus materials and testing, particularly Daphne Jansen, Jefta Lagerwerf, Inge Stok and Sam Theunissen. This work was funded by the Max Planck Society.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The dataset generated and analysed during the current study will be made available on https://archive.mpi.nl/mpi/ for academic researchers upon publication of all articles based on the current study.

ETHICS APPROVAL STATEMENT

The study was approved by the Ethical Board of Social Sciences, Radboud University, Nijmegen, and conforms to the standards of the Declaration of Helsinki.

ORCID

Melis Çetinçelik https://orcid.org/0000-0002-8931-5732

REFERENCES

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). CELEX2 [Data set]. Linguistic Data Consortium. https://doi.org/10.35111/GS6S-GM48

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Basirat, A., Brunellière, A., & Hartsuiker, R. (2018). The role of audiovisual speech in the early stages of lexical processing as revealed by the ERP word repetition effect: Audiovisual speech in early lexical processing. *Language Learning*, *68*, 80–101. https://doi.org/10.1111/lang.12265

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Bergmann, C., & Cristia, A. (2016). Development of infants' segmentation of words from native speech: A meta-analytic approach. *Developmental Science*, *19*(6), 901–917. https://doi.org/10.1111/desc.12341

Boersma, P., & Weenink, D. (2021). Praat: Doing Phonetics by Computer [Computer program]. http://www.praat.org/

Brent, M., & Siskind, J. (2001). The role of exposure to isolated words in early vocabulary. *Cognition*, *81*, B33–44. https://doi.org/10.1016/S0010-0277(01)00122-6

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207–220. https://doi.org/10.1017/S030500090700829X

Brown, V. A., & Strand, J. F. (2019). About face: Seeing the talker improves spoken word recognition but increases listening effort. *Journal of Cognition*, *2*(1), 1. https://doi.org/10.5334/joc.89

Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2021). Do the eyes have it? A systematic review on the role of eye gaze in infant language development. *Frontiers in Psychology*, *11*, https://doi.org/10.3389/fpsyg.2020.589096

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

D'Entremont, B. (2000). A perceptual-attentional explanation of gaze following in 3-and 6-month-olds. *Developmental Science*, *3*(3), 302–311. https://doi.org/10.1111/1467-7687.00124

ELAN. (2022). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive (Version 6.3). https://archive.mpi.nl/tla/elan

Farroni, T., Menon, E., & Johnson, M. H. (2006). Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child Psychology*, *95*(4), 298–308. Scopus. https://doi.org/10.1016/j.jecp.2006.08.001

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego, CA: Singular Publishing Group.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/s0305000900010679

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2013). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes*, *28*, 1207–1223. https://doi.org/10.1080/01690965.2012.701758

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). "Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations": Erratum. *Infancy*, *10*(2), 199. https://doi.org/10.1207/s15327078in1002_6

Goyet, L., de Schonen, S., & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research*, *1332*, 75–89. https://doi.org/10.1016/j.brainres.2010.03.047

Graham, S. A., Nilsen, E. S., & Nayer, S. L. (2007). Following the intentional eye: The role of gaze cues in early word learning. In *Gaze-following: Its development and significance. Visual perception* (Vol. 2323, pp. 193–216). Psychology Press.

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*(3), 692–697. https://doi.org/10.1016/j.cognition.2014.08.008

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, *76*(3), 598–613. https://doi.org/10.1111/j.1467-8624.2005.00866.x

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, *7*(3), 504–509. https://doi.org/10.3758/BF03214363

Junge, C., Cutler, A., & Hagoort, P. (2014). Successful word recognition by 10-Month-olds given continuous speech both at initial exposure and test. *Infancy*, *19*(2), 179–193. https://doi.org/10.1111/infa.12040

Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science*, *15*(4), 463–473. https://doi.org/10.1111/j.1467-7687.2012.1144.x

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. https://doi.org/10.1006/cogp.1995.1010

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3–4), 159–207. https://doi.org/10.1006/cogp.1999.0716

Kidd, E., Junge, C., Spokes, T., Morrison, L., & Cutler, A. (2018). Individual differences in infant speech segmentation: Achieving the lexical shift. *Infancy*, *23*(6), 770–794. https://doi.org/10.1111/infa.12256

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Brain Research. Cognitive Brain Research*, *24*(1), 109–116. https://doi.org/10.1016/j.cogbrainres.2004.12.009

Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from dutch ten-month-olds. *Infancy*, *14*(6), 591–612. https://doi.org/10.1080/15250000903263957

Kooijman, V., Johnson, E. K., & Cutler, A. (2008). Reflections on reflections of infant word recognition. In A. D. Friederici, & G. Thierry (Eds.), *Trends in language acquisition research* (Vol. 5, pp. 91–114). John Benjamins Publishing Company. https://doi.org/10.1075/tilar.5.06koo

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., & Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology*, *4*, https://doi.org/10.3389/fpsyg.2013.00025

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lalonde, K., & Werner, L. A. (2021). Development of the mechanisms underlying audiovisual speech perception benefit. *Brain Sciences*, *11*(1), 1. https://doi.org/10.3390/brainsci11010049

Lavelli, M., & Fogel, A. (2005). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: The 2-month transition. *Developmental Psychology*, *41*(1), 265–280. https://doi.org/10.1037/0012-1649.41.1.265

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

Lloyd-Fox, S., Széplaki-Köllőd, B., Yin, J., & Csibra, G. (2015). Are you talking to me? Neural activations in 6-month-old infants in response to being addressed during natural interactions. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, *70*, 35–48. https://doi.org/10.1016/j.cortex.2015.02.005

LoBue, V., Buss, K. A., Taber-Thomas, B. C., & Pérez-Edgar, K. (2017). Developmental differences in infants' attention to social and nonsocial threats. *Infancy*, *22*(3), 403–415. https://doi.org/10.1111/infa.12167

Lusk, L. G., & Mitchel, A. D. (2016). Differential gaze patterns on eyes and mouth during audiovisual speech segmentation. *Frontiers in Psychology*, *7*, https://doi.org/10.3389/fpsyg.2016.00052

Makeig, S., Bell, A. J., Jung, T., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In Advances in Neural Information Processing Systems (pp. 145–151).

Männel, C., & Friederici, A. D. (2013). Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex*, *49*(10), 2788–2798. https://doi.org/10.1016/j.cortex.2013.09.003

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

Mitchel, A. D., & Weiss, D. J. (2014). Visual speech segmentation: Using facial cues to locate word boundaries in continuous speech. *Language, Cognition and Neuroscience*, *29*(7), 771–780. https://doi.org/10.1080/01690965.2013.791703

Moradi, S., Lidestam, B., & Rönnberg, J. (2013). Gated audiovisual speech identification in silence vs. noise: Effects on time and accuracy. *Frontiers in Psychology*, *4*, 350.

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283–298. https://doi.org/10.1016/S0193-3973%2899%2900040-4

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Neal, R., & Schwartz, H. K. (2000). Gaze following, temperament, and language development in 6-month-olds: A replication and extension. *Infant Behavior & Development*, *23*(2), 231–236. https://doi.org/10.1016/S0163-6383(01)00038-8

Nazzi, T., Dilley, L. C., Jusczyk, A. M., Shattuck-Hufnagel, S., & Jusczyk, P. W. (2005). English-learning infants' segmentation of verbs from fluent speech. *Language and Speech*, *48*, 279–298. https://doi.org/10.1177/00238309050480030201

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*, 156869. https://doi.org/10.1155/2011/156869

Parise, E., Handl, A., Palumbo, L., & Friederici, A. D. (2011). Influence of eye gaze on spoken word processing: An ERP study with infants. *Child Development*, *82*(3), 842–853. https://doi.org/10.1111/j.1467-8624.2010.01573.x

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*(2), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

Polka, L., Rvachew, S., & Molnar, M. (2008). Speech perception by 6- to 8-month-olds in the presence of distracting sounds. *Infancy*, *13*(5), 421–439. https://doi.org/10.1080/15250000802329297

R Core Team. (2022). R: A language and environment for statistical computing *(4.2.2)*. R Foundation for Statistical Computing. http://www.R-project.org/

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8 month old infants. *Science*, https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*(3), 587–597. https://doi.org/10.1016/j.neuropsychologia.2006.03.036

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Snijders, T. M., Benders, T., & Fikkert, P. (2020). Infants segment words from songs—an EEG study. *Brain Sciences*, *10*(1), 39. https://doi.org/10.3390/brainsci10010039

Snijders, T. M., Kooijman, V., Cutler, A., & Hagoort, P. (2007). Neurophysiological evidence of delayed segmentation in a foreign language. *Brain Research*, *1178*, 106–113. https://doi.org/10.1016/j.brainres.2007.07.080

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Tan, S. H. J., Kalashnikova, M., Di Liberto, G. M., Crosse, M. J., & Burnham, D. (2022). Seeing a talking face matters: The relationship between cortical tracking of continuous auditory-visual speech and gaze behaviour in infants, children and adults. *NeuroImage*, *256*, 119217. https://doi.org/10.1016/j.neuroimage.2022.119217

Taylor, G., & Herbert, J. S. (2013). Eye tracking infants: Investigating the role of attention during learning on recognition memory. *Scandinavian Journal of Psychology*, *54*(1), 14–19. https://doi.org/10.1111/sjop.12002

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development*. *Journal of Child Language*, *42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–B34. https://doi.org/10.1016/j.cognition.2005.01.006

Van De Weijer, J. (1999). Language input for word discovery [PhD thesis, Radboud University] Application/pdf]. https://doi.org/10.17617/2.2057670

van der Velde, B., White, T., & Kemner, C. (2021). The emergence of a theta social brain network during infancy. *NeuroImage*, *240*, 118298. https://doi.org/10.1016/j.neuroimage.2021.118298

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Von Holzen, K., Nishibayashi, L.-L., & Nazzi, T. (2018). Consonant and vowel processing in word form segmentation: An infant ERP study. *Brain Sciences*, *8*(2), 2. https://doi.org/10.3390/brainsci8020024

Wass, S. V., Whitehorn, M., Marriott Haresign, I., Phillips, E., & Leong, V. (2020). Interpersonal neural entrainment during early social interaction. *Trends in Cognitive Sciences*, *24*(4), 329–342. https://doi.org/10.1016/j.tics.2020.01.006

Zink, I., & Lejaegere, M. (2002). N-CDI's: lijsten voor communicatieve ontwikkeling, aanpassing en hernormering van de MCArthur CDI's van Fenson et al. Acco.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2024). Does the speaker's eye gaze facilitate infants' word segmentation from continuous speech? An ERP study. *Developmental Science*, *27*, e13436. https://doi.org/10.1111/desc.13436