Journal Pre-proof


Natural Infant-Directed Speech Facilitates Neural Tracking of Prosody

Katharina H. Menn, Christine Michel, Lars Meyer, Stefanie Hoehl, Claudia Männel

PII: S1053-8119(22)00120-3
DOI: https://doi.org/10.1016/j.neuroimage.2022.118991
Reference: YNIMG 118991

To appear in: *NeuroImage*

Received date: 23 September 2021
Revised date: 2 February 2022
Accepted date: 10 February 2022

Please cite this article as: Katharina H. Menn, Christine Michel, Lars Meyer, Stefanie Hoehl, Claudia Männel, Natural Infant-Directed Speech Facilitates Neural Tracking of Prosody, *NeuroImage* (2022), doi: https://doi.org/10.1016/j.neuroimage.2022.118991

## Highlights

- We investigate infantsâĂŹ tracking of natural infant- and adult-directed speech
- Mothers enhance prosodic stress in infant-directed speech
- Infants track the prosodic stress and syllable rate for natural speech
- Infant-directed speech facilitates infantsâĂŹ tracking of prosodic stress

# Natural Infant-Directed Speech Facilitates Neural Tracking of Prosody

Katharina H. Menn*[1,2,3], Christine Michel*[4,5], Lars Meyer[2,6], Stefanie HoehlâĂă[7], and Claudia MÃďnnelâĂă[1,8]

[1]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany

[2]Research Group Language Cycles, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany

[3]International Max Planck Research School on Neuroscience of Communication: Function, Structure, and Plasticity, Stephanstr 1a, 04103 Leipzig, Germany

[4]Research Group Early Social Cognition, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany

[5]Faculty for Education, Leipzig University, MarschnerstraÃ§e 31, 04109 Leipzig, Germany

[6]Clinic for Phoniatrics and Pedaudiology, University Hospital MÃijnster, Albert-Schweitzer-Campus 1, 48149 MÃijnster, Germany

[7]University of Vienna, Faculty of Psychology, UniversitÃďtsring 1, 1010 Vienna, Austria

[8]Department of Audiology and Phoniatrics, CharitÃľ âĂŞ UniversitÃďtsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany

*,âĂă These authors each contributed equally

**Author note:** Correspondence concerning this article should be addressed to Katharina H. Menn, E-mail: menn@cbs.mpg.de

## Abstract

Infants prefer to be addressed with infant-directed speech (IDS). IDS benefits language acquisition through amplified low-frequency amplitude modulations. It has been reported that this amplification increases electrophysiological tracking of IDS compared to adult-directed speech (ADS). It is still unknown which particular frequency band triggers this effect. Here, we compare tracking at the rates of syllables and prosodic stress, which are both critical to word segmentation and recognition. In mother-infant dyads (n=30), mothers described novel objects to their 9-month-olds while infants' EEG was recorded. For IDS, mothers were instructed to speak to their children as they typically do, while for ADS, mothers described the objects as if speaking with an adult. Phonetic analyses confirmed that pitch features were more prototypically infant-directed in the IDS-condition compared to the ADS-condition. Neural tracking of speech was assessed by speech-brain coherence, which measures the synchronization between speech envelope and EEG. Results revealed significant speech-brain coherence at both syllabic and prosodic stress rates, indicating that infants track speech in IDS and ADS at both rates. We found significantly higher speech-brain coherence for IDS compared to ADS in the prosodic stress rate but not the syllabic rate. This indicates that the IDS benefit arises primarily from enhanced prosodic stress. Thus, neural tracking is sensitive to parentsâĂŹ speech adaptations during natural interactions, possibly facilitating higher-level inferential processes such as word segmentation from continuous speech.

*Keywords:* EEG, speech-brain coherence, speech entrainment, infant-directed speech, natural interaction, adult-directed speech

Natural Infant-Directed Speech Facilitates Neural Tracking of Prosody

## 1. Introduction

Across many languages, adults address infants in a characteristic register termed infant-directed speech (IDS) (Soderstrom, 2007; Cristia, 2013; Fernald et al., 1989). IDS differs from adult-directed speech (ADS) along acoustic and linguistic dimensions. In particular, IDS contains exaggerated prosodic cues (Fernald et al., 1989; Grieser and Kuhl, 1988; Fernald and Simon, 1984; Katz et al., 1996), is syntactically simpler (Soderstrom et al., 2008) and may be spoken more slowly (Raneri et al., 2020) with expanded vowel sounds (Green et al., 2010; Adriaans and Swingley, 2017). Previous electrophysiological work has indicated that these IDS characteristics benefit infants' speech processing (e.g. Háden et al., 2020; Zangl and Mills, 2007). While earlier EEG studies mostly focused on event-related potentials, we here employ EEG to examine infants' online speech processing continuously. There are indications that IDS benefits infants' language acquisition in particular. Frequent exposure to IDS boosts later vocabulary development (RamÃŋrez-Esparza et al., 2014; Weisleder and Fernald, 2013) and laboratory studies showed that IDS assists infants' word segmentation (Schreiner and Mani, 2017; Thiessen et al., 2005) and recognition (Singh et al., 2009; Männel and Friederici, 2013), and their acquisition of word-object associations (Graf Estes and Hurley, 2013) over ADS.

Which specific acoustic cues in IDS help infants' language acquisition? Candidates include increased fundamental frequency (F0) and F0 modulation (see Spinelli et al., 2017, for a meta-analysis). In recent years, a particular focus has been put on the amplitude modulation structure in IDS. Continuous speech contains acoustic information at different timescales, which to a certain extend correspond to linguistic units, such as phonemes, syllables, and intonation phrases. In particular, the amplitude envelope conveys the boundaries of linguistic units even to infant listeners who lack vocabulary as such (see also Goswami, 2019). Leong and Goswami (2015) analyzed the amplitude modulation structure of nursery rhymes, a particularly rhythmic form of IDS, which were read by female speakers prompted with a picture depicting young children. The authors found that amplitude modulations are centered around three frequency rates, which match the occurrence rates of: prosodic stress (~2Hz), syllables (~5Hz), and phonemes (~20 Hz). When comparing spontaneously produced IDS during mother-infant interactions to ADS that the mother produced when interacting with another adult, Leong et al. (2017) found that amplitude modulations of prosodic stress are enhanced for IDS compared to ADS. This exaggeration of prosodic stress in IDS may be beneficial for infants' language development, as stress can provide an important cue for word onsets in naturalistic speech (Stärk et al., 2021; Jusczyk et al., 1999; Cutler and Carter, 1987) and thus aid word segmentation. If infants are sensitive to the pronounced stress modulations in IDS, these could thus provide an important stepping stone into language acquisition.

Recent studies have shown that infants' neural activity tracks speech by synchronizing with amplitude modulations corresponding to prosodic stress and syllables in nursery rhymes (Attaheri et al., 2021). For adults, it has been shown that the synchronization between neural activity and speech acoustics supports the segmentation and identification of linguistic units in speech (see Meyer, 2018) and relates to better language comprehension (Peelle et al., 2013; Doelling et al., 2014). Importantly, infants were shown to start tracking simple repeated sentences from birth (Ortiz Barajas et al., 2021). This early emergence suggests that neural tracking may support language development by aligning neural activity with speech-relevant amplitude modulations. At least by 7-months of age, infants' tracking is sensitive to the kind of speech register (IDS vs. ADS) and IDS benefits tracking of speech over ADS (Kalashnikova et al., 2018). It remains unclear, however, whether this benefit results specifically from prosodic stress or other speech characteristics, such as the syllable rhythm.

We here assess infants' tracking of speech in a naturalistic mother-infant interaction. The use of naturalistic IDS has the benefit of high ecological validity, as it elucidates infants' neural processing of the speech input they typically receive and thus increases generalizability of findings. Naturalistic stimuli allow for the dissociation of multiple levels of information in parallel (see also Jessen et al., 2021). For this reason, the number of studies relying on naturalistic input for investigating infants' neural processing of speech has recently started to increase and stimuli included recordings taken from natural mother-infant interactions (Kalashnikova et al., 2018), TV cartoons (Jessen et al., 2019) and one study even directly assessed face-to-face interactions (Lloyd-Fox et al., 2015). In face-to-face interactions, the speaker's visual cues are contingent with infant responses, which is difficult to manipulate in classical experiments. For the current study, the most relevant of these contingent cues is eye contact between parents and infants (mutual gaze), which was shown to increase neural processing of speech if combined with IDS (Lloyd-Fox et al., 2015). However, given the difficulty of manipulating mutual gaze experimentally, the specific effects on infants' speech processing are currently not well understood (for a review, see Çetinçelik et al., 2020).

In the current study we focus on the association between parental acoustic speech adaptations and infants' tracking, aiming at delineating whether neural tracking is facilitated by prosodic stress (defined by pitch contours) or syllable information (defined by the mean syllable duration) in IDS. To this end, we here contrast 9-month-old infants' responses to their mothers' IDS versus ADS at the stress rate and the syllabic rate. Focusing on 9-month-olds is particularly interesting, as infants at this age have started segmenting words from continuous speech but still mostly rely on prosodic cues (Schreiner and Mani, 2017; Männel and Friederici, 2013), meaning that information in the prosodic stress rate is particularly relevant for their word segmentation (Kooijman et al., 2009). In mother-infant dyads, mothers described novel objects to their 9-month-olds while the infants' electroencephalogram (EEG) was recorded. For IDS, parents were instructed to speak to their infants as they typically do, while for ADS, parents were supposed to describe the objects pretending they talk to an adult without looking at the infant or calling their name. Infants' tracking of maternal speech during the interactions was assessed using speech-brain coherence, which measures the synchronization between the neural signal and the speech envelope. We hypothesized that infants show speech-brain coherence at both the stress rate and the syllable rate. Concerning the difference between IDS and ADS processing, we postulate that IDS facilitates tracking (Kalashnikova et al., 2018) and that this facilitation is driven by enhanced amplitude modulations of prosodic stress (Leong et al., 2017).

## 2. Method

The present study reanalyzed data from a previous experiment, which assessed the influence of ostensive cues on infants' visual object encoding (Michel et al., 2021). Parents were asked to show and describe a total of 12 novel objects to their infant during a familiarization phase. Half of the objects were described naturally (IDS-condition), the other half were described without ostensive cues (i.e., mutual gaze, calling the infant by their name, and infant-directed speech; ADS-condition). Importantly, parents were asked to refrain from naming the objects. Given the aim of the present study to examine infants' neural processing of natural parental speech, we here assessed infants' tracking of maternal speech during the mother-infant interactions. Only the object description phase was analyzed for the purpose of the current study and will be described in this manuscript.

### 2.1 Participants

The final participant sample consisted of 30 German-learning infants (22 female) and their mothers. On average, infants were 9 months 12 days old (range: 9 months 0 days - 9 months, 29 days). Infants were born full-term (> 37 weeks), healthy, and raised in monolingual German environments. Our sample size was determined by the previously collected dataset. Michel et al. (2021) based their sample size on studies investigating infants' object encoding using similar paradigms and measures (e.g. Hoehl et al., 2014; Begus et al., 2015).

Additional 51 mother-infant (16 female, $M_{age} = 9$ months 15 days) interactions were tested, but not included in the current analysis due to less than 30 s total maternal speech in one of the conditions ($n = 17$), more than 4 noisy electrodes ($n = 1$), failure to reach the minimum criterion of 20 EEG epochs per condition after artifact rejection ($n = 19$), premature birth ($n = 1$), technical error ($n = 6$), or infant fussiness ($n = 7$). Because of the different foci of this manuscript and the original study (Michel et al., 2021), the exclusion criteria differed between the manuscripts and only 19 infants were commonly included in both. Informed written consent was obtained from the mothers before the experiment and ethical approval for the experimental procedure and reanalysis of the data was obtained from the Medical Faculty of the University of Leipzig. All work was conducted in accordance with the Declaration of Helsinki. The conditions of our ethics approval do not permit public archiving of participant data. Readers seeking access to the data should contact the corresponding author to arrange a formal data sharing agreement.

## 2.2 Procedure

Mothers and infants were seated across a small table. Infants sat in a baby chair while their electrophysiological activity was continuously recorded using EEG. Mother-infant interactions were recorded on video using four cameras and maternal speech was recorded using a microphone that was placed on the table in front of the mother (see Figure 1A).

The study consisted of 4 blocks, during each of which the mother held three novel objects above the table and spoke about them to her infant. The blocks alternated between the IDS-condition and the ADS-condition. The only difference between the two conditions was the way in which the mother was asked to describe the objects. Mothers were told that the aim of the study was to investigate the difference between joint observation and individual processing of objects on infants' visual object encoding, as this was the goal of the original study. They were specifically told to focus on eye gaze and

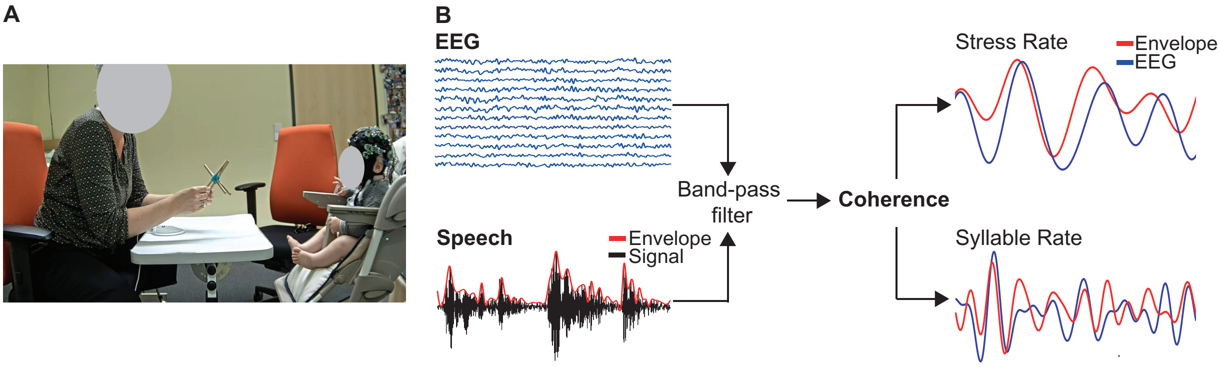


*Figure 1*. Overview of the experiment and analysis. **(A)** Example of the setting during the mother-infant interactions. Mother and infant sat across each other at a table. The mother held a novel object and described it to her infant either using IDS or using ADS, while the infant's EEG was recorded. **(B)** Overview of the speech-brain coherence analysis. Cleaned EEG and speech envelope were band-pass filtered in two frequency bands: prosodic stress rate and syllable rate. Coherence between EEG and envelope was computed for each electrode in both frequency bands.

speech. In the IDS-condition, the mother was asked to speak to her infant as she normally would when interacting with a novel object. She was specifically told that she could use IDS, call their infant's name and look at the infant. In the ADS-condition, the mother was instructed to describe the object as if she were speaking to an adult, that is she was asked to imagine that she was talking to herself or describing the objects to a close friend. She was also asked to refrain from calling the infant's name and looking at the infant, and specifically from establishing eye gaze during the ADS-condition. In both conditions, the infant was not allowed to touch the objects. The condition of the first block was counterbalanced between dyads. Mothers were given standardized oral and written instructions and were reminded of the procedure before every block.

Each block started with a 20 s baseline, during which infant and mother looked at soap bubbles produced by an experimenter. Afterwards, the object description phase started either after mutual gaze between infant and parent had been established (IDS-condition) or after the child looked at the mother (ADS-condition). In both conditions, the trial ended after the infant looked at the object for a cumulative total of 20 s. Looking duration was coded online by an experimenter observing the interactions on a screen. A second experimenter then announced the end of a trial by thanking the mother and switched the object. Average trial duration was 39.2 s ($SD$ = 8.6; see Supplementary Figure 1 for an overview of the whole procedure). Mothers were unaware of the looking time criterion. None of the objects had eyes or face-like features on it. Pretests with an independent sample of infants confirmed that, in general, infants were unfamiliar with the objects and all objects were similarly interesting to infants.

## 2.3 Speech Processing

**2.3.1 Preprocessing.** Audio recordings were annotated and analyzed using Praat (Boersma, 2001). We annotated every instance of maternal speech during the object description phase, excluding fragments with any non-speech interference. Instances of such interference included: infant vocalizations, laughter, external noise, or (rhythmic) non-speech sounds, such as knocking the object on the table, scratching the surface of the object or tapping against the object. Speech segments with pauses longer than 1000 ms were coded as separate segments.

**2.3.2 Amplitude Envelope.** The broad-band amplitude envelope of the audio signals was computed following Gross et al. (2013) using the Chimera toolbox (Smith et al., 2002). The intensity of the speech signal was normalized per condition. We divided the frequency spectrum from 100 - 8000 Hz into nine frequency bands equally spaced on the cochlea. The audio signal was band-pass filtered into these frequency bands with a fourth order Butterworth filter (forward and backward). Afterwards, the absolute values of the Hilbert transform were computed for each band and averaged across bands. Last, the envelope was downsampled to 500 Hz, which corresponds to the sampling rate of the EEG signal.

In addition, we computed the pitch envelope for both conditions separately. For this we determined the respective F0 range for both speech conditions (IDS: 145 - 392 Hz; ADS: 138 - 325 Hz), which we divided into three frequency bands equally spaced on the cochlea. We then followed the same procedure as described for the broad-band envelope.

**2.3.3 Frequency Bands.** To identify the syllable rate of mothers' IDS and ADS, we annotated the duration of all syllables for the dyads included in the final analysis. The average syllable duration was 194 ms for the ADS-condition and 181 ms for the IDS-condition. The syllable rate was determined as the 2 Hz window centered around the average syllable duration (ADS: 194 ms or 5.15 Hz; IDS: 181 ms or 5.5 Hz), leading to 4.15 Hz - 6.15 Hz for ADS and 4.5 - 6.5 Hz for IDS.

The prosodic stress rate of mothers' speech was identified based on the pitch envelope. For this, we segmented the parts of the pitch envelope corresponding to uninterrupted maternal speech into epochs of 2 s length with 50% overlap. We then computed the Fourier transform of each epoch using Slepian multitapers and averaged the resulting power spectral density (PSD) estimate across epochs and dyads for both speech conditions. The averaged PSD was visually inspected for deviations from the aperiodic 1/f noise. This way the frequency band for the prosodic stress rate was determined as 1 - 2.5 Hz. We decided not to assess amplitudes below 1 Hz since this is the high-pass frequency recommended for the preprocessing of developmental EEG data (see e.g. Gabard-Durnam et al., 2018). The bands identified for the prosodic stress rate and the syllable rate were in line with rates reported in previous studies (e.g. Leong and Goswami, 2015; Chandrasekaran et al., 2009).

**2.3.4 Amplitude Modulations.** To compute the amplitude modulations at the syllable rate, we filtered the broad-band amplitude envelope in the corresponding frequency bands for IDS and ADS. We then segmented the parts of the envelope corresponding to uninterrupted maternal speech into epochs of 2 s length with 50%

overlap. Root mean square values were computed for every epoch and averaged across epochs for both speech conditions.

Amplitude modulations in the prosodic stress rate were computed based on the pitch envelope. We band-pass filtered the pitch envelope in the frequency band corresponding to prosodic stress before proceeding in the same way as described for the syllable rate.

### 2.4 Experimental Manipulation Check

To assess whether the speech in the IDS-condition was more typically infant-directed than speech in the ADS-condition, we measured the mean F0 and F0 range (between the 5th and the 95th percentile) of maternal speech in both conditions as an acoustic correlate of IDS (see, Spinelli et al., 2017). In addition, we tested whether the amplitude modulations in the prosodic stress rate and the syllable rate differed between IDS versus ADS. We ran separate t-tests for each acoustic measure, assessing a difference between the IDS- and the ADS-condition. Note that we opted for separate tests in assessing condition differences in amplitude modulations in the two frequency bands since they were computed based on different envelopes and are therefore not directly comparable. Resulting p-values were corrected for multiple comparisons using false discovery rate (FDR-correction).

### 2.5 EEG-Recording and Preprocessing

EEG was recorded with a 32-channel EasyCap system by Brain Products GmbH, with active electrodes arranged according to the 10/10 system. The sampling rate of the recordings was 500 Hz. The right mastoid served as the online reference and vertical electrooculograms were recorded bipolarly if tolerated by the infant.

EEG processing was done using the publicly available 'eeglab' (Delorme and Makeig, 2004) and 'fieldtrip' (Oostenveld et al., 2011) toolboxes as well as custom Matlab code (The MathWorks, Inc., Natick, US). EEG preprocessing was done automatically using a modified version of the Harvard Automated Preprocessing Pipeline (HAPPE: Gabard-Durnam et al., 2018). In line with HAPPE, data was re-referenced to Cz to obtain symmetrical components in the ICA, high-pass filtered with a noncausal finite impulse response filter (pass-band: 1 Hz, -6 dB cutoff: 0.5 Hz) and electrical line noise (50 Hz) was removed using ZapLine from NoiseTools (de Cheveigné, 2020). Noisy channels were identified by assessing the normed joint probability of the average log power from 1 - 125 Hz and rejected if exceeding a threshold of 3 SD from the mean (mean number of removed channels = 1; range: 0-4). We applied a wavelet-enhanced ICA (Castellanos and Makarov, 2006) with a threshold of 3 to remove large artifacts, before the data was decomposed with ICA and artifact-related components were automatically rejected using MARA (Winkler et al., 2011, ;mean number of rejected components = 14, range: 7-25). Afterwards, noisy channels were interpolated using spherical splines and the data was re-referenced to the linked mastoids.

EEG data and the broad-band speech envelope were band-pass filtered at the stress and syllable rate. Filter order was optimised through the ParksâĂŞMcLellan algorithm (Parks and McClellan, 1972). For the prosodic stress band, this resulted in a 14572th-order one-pass 1âĂŞ2.5-Hz band-pass filter. The phase shift was compensated for by an according time shift. For the syllabic band, we used an 15883th-order one-pass filter with pass-frequencies of 4.5 - 6.5 Hz for IDS and 4.15 - 6.15 Hz for ADS. All data were padded before filter application.



The artifact-corrected EEG data was segmented into continuous trials corresponding to the annotated maternal speech and combined with the respective broad-band speech envelope, which had been downsampled to 500 Hz. The combined data was segmented into 2 second epochs with 50% overlap. Epochs with amplitudes exceeding $\pm 40\mu$V in any channel were rejected automatically. On average, infants contributed a total of 112 epochs to the analysis ($M_{IDS} = 57.8$, $SD = 27.4$; $M_{ADS} = 54.2$, $SD = 32.8$). The 23 channels included in the final analysis were: Fz, F3/4, F7/8, FC1/2, FC3/4, FT7/8, Cz, C3/4, T7/8, CP3/4, Pz, P3/4, and P7/8. We removed the outer channels from the final analysis, since the EEG signal was consistently noisy across infants.

## 2.6 Data Analysis

**2.6.1 Speech-Brain Coherence.** The relationship between speech and brain signal was quantified using Hilbert coherence over time (see Figure 1B). The coherence value measures the phase-synchronization between the EEG signal and the corresponding speech envelope, weighted by their relative amplitude. Coherence is measured on a scale from 0 (random coupling) to 1 (perfect synchronization).

Coherence between speech envelope and individual electrodes in both frequency rates was computed according to the formula: $Coh_{xy}(f) = \frac{|P_{xy}(f)^2|}{P_{xx}(f)P_{yy}(f)}$, where $P_{xy}(f)$ is the cross-spectral density between the band-pass filtered speech and EEG signal, and $P_{xx}(f)$ and $P_{yy}(f)$ are the auto-spectral density of the speech and EEG signal, respectively.

To analyze whether speech-brain coherence was higher than expected by chance, the observed coherence values were compared against surrogate data. Surrogate data was created by randomly pairing the epoched EEG data with the broad-band speech envelope from a randomly selected epoch from the same or a different dyad and applying a circular shift to the envelope time series (Keitel et al., 2017). This process was repeated for 10,000 permutations.

**2.6.2 Analyses.** The observed and permuted coherence values for each infant were averaged across trials and channels. P-values were derived as the proportion of coherence values in the permutation distribution exceeding the observed value. To assess differences between IDS and ADS, we ran a repeated-measures ANOVA with speech condition (IDS vs. ADS) and frequency rate (syllabic rate vs. prosodic rate) as within-subjects factors.

Table 1

*Analysis of speech acoustics. Standard deviation in brackets*

| Acoustic Measure | | IDS | ADS | p-value |
|---|---|---|---|---|
| Pitch (F0) | Mean | 238 Hz (28) | 214 Hz (19) | < .001 |
| | Range | 247 Hz (62) | 188 Hz (49) | < .001 |
| Amplitude Modulations (a.u.; $1{\times}10^{-3}$) | Stress Rate | 2.5 (0.50) | 2.1 (0.46) | < .001 |
| | Syllable Rate | 1 (0.14) | 0.96 (0.15) | .482 |

## 3. Results

Maternal speech in the IDS-condition was more prototypically infant-directed than in the ADS-condition. Speech had a significantly higher mean pitch, $t(29) = 7.2$, $p < .001$, and pitch range, $t(29) = 6.21$, $p < .001$, in the IDS-condition compared to the ADS-condition. The amplitude modulations were significantly higher for IDS than ADS in the stress rate, $t(29) = 4.1$, $p < .001$, but not in the syllable rate, $t(29) = 0.71$, $p = .482$. Table 1 summarizes the descriptive statistics of the acoustic measures. For further summary statistics of speech content, see supplementary Table 1.

The permutation test showed significant speech-brain coherence for both the prosodic stress rate, $p < .001$, and the syllable rate, $p < .001$ (Figure 2A). The repeated-measures ANOVA showed a significant main effect of speech condition, $F(1, 29) = 160.77$, $p < .001$, and no significant main effect of frequency rate, $F(1, 29) = 2.43$, $p = .13$ . Importantly, we observed a significant interaction between speech condition and frequency rate, $F(1, 29) = 9.14$, $p = .005$ (Figure 2B). Follow-up t-tests revealed that speech-brain coherence for the stress rate was significantly higher in the IDS-condition ($M_{IDS} = 0.492$, $SD = 0.025$) than in the ADS-condition ($M_{ADS} = 0.476$, $SD = 0.022$), $t(29) = 3.4$, $p = .002$. We found no evidence for a difference between the IDS-condition ($M_{IDS} = 0.42$, $SD = 0.02$) and the ADS-condition ($M_{ADS} = 0.425$, $SD = 0.02$) for the syllable rate, $t(29) = -0.99$, $p = .33$. Analyses were repeated on

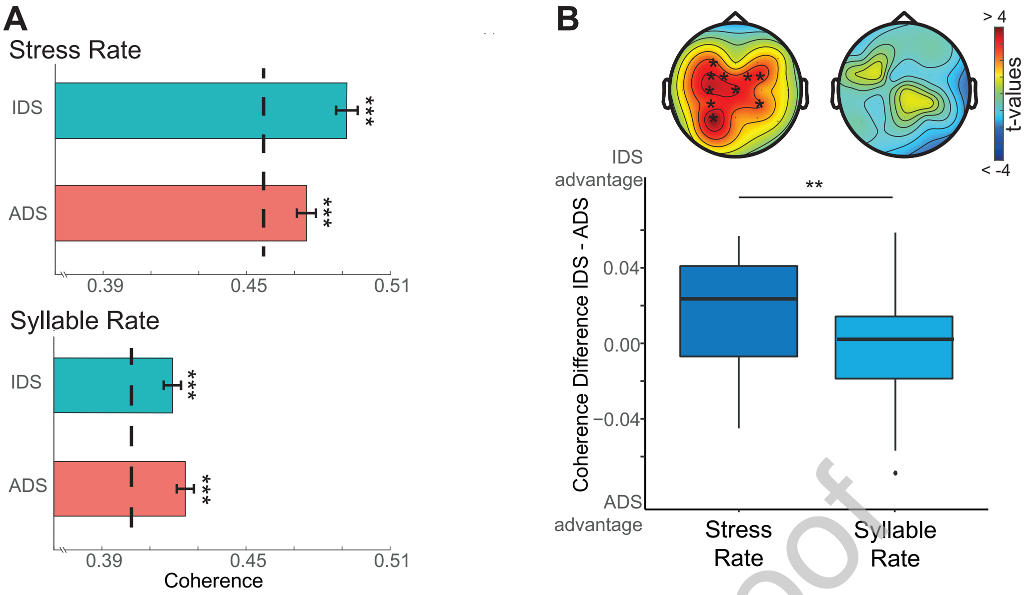


*Figure 2*. Overview of our results. **(A)** Coherence values were averaged across all electrodes. Errorbars depict standard errors. Dashed lines indicate 95% significance cut-offs based on a permutation baseline. Speech-brain coherence was significantly higher than chance for both IDS and ADS in the two frequency rates. **(B)** Scalp topography for the comparison IDS versus ADS. Asteriscs indicate electrodes included in the cluster in the control analysis. For the main analysis, we compared averages across all electrodes. The difference between IDS and ADS was significantly higher in the stress rate than in the syllable rate.

non-normalized data to ensure that the difference between conditions did not arise from intensity differences. The pattern of the results did not change.

## 3.1 Control Analysis: Ostensive Cues

Ostensive cues potentially influence speech processing (see Çetinçelik et al., 2020; Csibra and Gergely, 2009). In our study, such cues were primarily present in the IDS-condition. We therefore conducted additional analyses to control for the possibility that the tracking difference between IDS and ADS observed in our study was based on differences in ostensive cues, specifically focusing on mutual eye gaze, infant looks to the mother's face and mentioning the infant's name.

In every frame of the video recording, mother's and infant's gaze were coded as looking to the object, to the face of the interaction partner, to the environment or as non-codeable. The reliability of the codes was excellent (ICC for mothers = 0.994, ICC for infants = 0.987). Mutual gaze was defined as periods with simultaneous gaze on the other interaction partner. We then reanalyzed the data excluding all epochs containing mutual eye gaze. On average, infants contributed a total of 103 epochs to the follow-up analysis ($M_{IDS} = 49.4$, $SD = 23.2$; $M_{ADS} = 54.1$, $SD = 32.7$). A paired t-test comparing the speech conditions in the stress rate showed that speech-brain coherence was still significantly higher for the IDS-condition ($M_{IDS} = 0.489$, $SD = 0.023$) than the ADS-condition ($M_{ADS} = 0.475$, $SD = 0.022$) after controlling for the effect of mutual eye gaze, $t(29) = 2.87$, $p = .0075$. It is, however, possible that infants show a sustained effect of mutual gaze beyond the epoch. We therefore also excluded the 5 epochs succeeding mutual eye gaze. This also did not change the pattern of our results. Note that we were unable to exclude the whole object description trial in which mutual eye gaze occured, as this would have left us with too few epochs for a reliable comparison. In addition, we compared tracking of IDS in the prosodic stress rate between infants with high mutual gaze to infants with low mutual gaze, grouped by a median split of the number of epochs containing mutual gaze. The two groups did not significantly differ, $t(28) = 0.467$, $p = .64$.

To assess the possibility that the IDS advantage for tracking in the prosodic stress rate was driven by maternal visual cues other than mutual gaze, we excluded all epochs in which the infant looked at the mother's face, irrespective of whether there was mutual gaze or not. On average, infants contributed a total of 90.9 remaining trials to this follow-up analysis ($M_{IDS} = 45.1$, $SD = 23.3$; $M_{ADS} = 45.8$, $SD = 26.57$). Speech-brain coherence in the prosodic stress rate remained significantly higher for the IDS-condition ($M_{IDS} = 0.489$, $SD = 0.026$) than the ADS-condition ($M_{ADS} = 0.472$, $SD = 0.025$) after excluding these epochs in which infants were looking at their mother's face, $t(29) = 3.07$,

$p = .0046$.

Lastly, we assessed whether the amount of calling the infant's name in the IDS-condition drove the IDS facilitation in the stress rate. On average, mothers called their infant's name 3.9 times in the IDS-condition ($SD = 3.7$). We compared tracking in the stress rate between infants who experienced high calling of their name versus infants who experienced low calling of their name, which were grouped based on a median split (median = 3.5). There was no significant difference between the two name-calling groups, $t(28) = 0.70$, $p = .489$. Note that we only controlled for instances in which the infantsâĂŹ full name or an abbreviation of it was mentioned, but not for other potentially attention-evoking phrases that mothers commonly use in IDS. We therefore cannot fully rule out that the use of such phrases increased attention specifically in the IDS condition.

### 3.2 Control Analysis: Topography

All EEG analyses reported before were done on coherence values averaged across the 23 selected electrodes. This approach may hide topography differences between the IDS- and the ADS-condition in the two frequencies of interest. To assess this possibility, we conducted a control analysis on the electrode level, using threshold-free cluster-enhancement with 10,000 permutations for multiple comparison correction (height-weight = 2, extend-weight = 0.5; Smith and Nichols, 2009). In line with our earlier results, we found a significant difference between the IDS- and the ADS-condition in the prosodic stress rate ($p < .001$), but not in the syllable rate. The difference in the stress rate was driven by a left-central cluster that included electrodes F3, FC3, FC1, C3, CP3, P3, Cz, FC2, FC4, and CP4. These electrodes are marked by asterisks in the topography plot in Figure 2B.

### 3.3 Control Analysis: Pauses

IDS has been related to an increased number of pauses compared to ADS (Martin et al., 2016), which may form acoustic edges that can contribute to speech-brain coherence (Gross et al., 2013). In line with earlier findings, the IDS-condition (25 pauses/min, $SD = 11.3$) had a higher rate of pauses than the ADS-condition (17.3 pauses/min, $SD = 11.1$), $t(29) = 3.82$, $p < .001$. Pause durations did not differ between the two conditions ($M_{IDS} = 259$ ms, $SD = 75$; $M_{ADS} = 250$ ms, $SD = 78$), $t(29) = 0.63$, $p = .536$. To assess whether the increased number of pauses in IDS contributes to the IDS advantage for tracking, we compared phase-clustering from 1 to 8 Hz (in steps of 0.5 Hz) at word onsets following pauses and thus forming an acoustic edge to phase-clustering at word onsets within continuous speech. The analysis assessed phase-clustering starting -100 ms before word onset until 1 second after in steps of 10 ms for all electrodes individually, and number of word onsets contributing to the analysis were matched. Our analysis used cluster-based permutation for multiple comparison correction and showed no significant difference in phase-clustering between the two types of word onsets ($p > .1$). Next, we compared phase-clustering at pause offset between the IDS- and the ADS-condition using the same frequencies and time window. The cluster-based permutation analysis showed no significant difference in phase-clustering between the two conditions ($p > .1$), giving no evidence that infants' neural responses to pauses differed between IDS and ADS. At last, we compared tracking in the stress rate between infants with a higher rate of pauses versus infants with a lower rate of pauses, grouped based on a median split (median = 25.8). The two groups showed no significant differences in tracking, $t(29) = 0.69$, $p = .5$. While this does not exclude the possibility that pauses and associated acoustic edges increase speech-brain coherence, we find no evidence that they are the main driver of the IDS facilitation for tracking in the stress rate.

## 4. Discussion

The present study set out to investigate infants' neural tracking of natural IDS compared to ADS and to delineate whether the IDS facilitation is driven by prosodic stress. We observed significant tracking of speech at both the stress and the syllable rate during natural interactions of 9-month-olds with their mothers. Adding to previous findings, we report here that tracking is facilitated by IDS and that this effect is specific to the prosodic stress rate. This suggests that the IDS advantage for infants' tracking is specifically based on enhanced prosodic stress and not on the syllable rhythm. Our finding emphasizes the important role of IDS for infants' speech processing and possibly their language development.

At the age of 9 months, infants have started to segment words from continuous speech (Jusczyk et al., 1999; Männel and Friederici, 2013; Junge et al., 2014), facilitated by IDS (Schreiner and Mani, 2017). Speech segmentation is crucial for the acquisition of higher-level linguistic meaning and better word segmentation in infancy was shown to predict later vocabulary size (Junge et al., 2012) and syntactic skills (Kooijman et al., 2013). Since continuous speech contains no pauses between words, infants must rely on other acoustic cues to detect word boundaries. In stress-based languages like English or German, stressed syllables can provide a valuable cue for segmenting words from continuous speech (Jusczyk et al., 1999), as the majority of content words in these languages have word-initial stress (Cutler and Carter, 1987; Stärk et al., 2021). Our study shows that that not only do mothers enhance their amplitude modulations at the prosodic stress rate in IDS, but also infants do track this enhancement. This suggests that tracking might facilitate higher-level inferential processes such as word segmentation.

Because of the way this study was set-up, the IDS-condition included a number of additional ostensive cues that were not present in the ADS-condition. Most relevant are the addition of mutual gaze between mother and infant and calling of the infant's name, as mothers were specifically told to focus on these cues. In addition, it is possible that mothers increased other visual cues in the IDS-condition, as adults were shown to exaggerate facial expressions such as lip and head movements when addressing children (Swerts and Krahmer, 2010; Smith and Strader, 2014; Green et al., 2010), which we were unable to assess in the current study. These ostensive cues are special as they help guiding infants' attention to maternal speech (Csibra and Gergely, 2006, 2009) and consequently may have assisted to increase infants' speech processing (for a review, see Çetinçelik et al., 2020). However, we find that the IDS-condition specifically facilitated tracking in the prosodic stress rate and no evidence for an IDS facilitation in the syllable rate. This finding is not compatible with a general increase of attention to maternal speech by ostensive cues in the IDS-condition. In addition, our control analysis showed that the IDS benefit for tracking persists even after we excluded epochs with mutual eye gaze and that infants who experienced more calling of their name did not show a higher tracking of IDS in the prosodic stress rate than infants who experienced less calling of their name. These results do not imply that visual information is irrelevant for speech processing. Previous studies have shown that visual information increases tracking of speech in adults (Crosse et al., 2015; Bourguignon et al., 2020) and likely also in children (Power et al., 2012). As our design does not allow to investigate whether the frequency of visual exaggerations in the IDS-condition coincides with the prosodic stress rate, we conducted a control analysis excluding all epochs during which the infant looked at the mother. Even for the parts of the interactions in which the infants did not look at the mother, the IDS tracking advantage in the prosodic stress rate persisted. This supports our conclusion that the IDS benefit for speech processing results from its acoustic properties, even though we cannot fully exclude the possibility that infants still perceived some exaggerated visual cues even if they did not directly look at the mother's face. Further studies are needed to dissociate the unique contributions of visual and acoustic cues to infants' neural processing of IDS.

Regarding parental acoustic speech modulations, the enhanced amplitude modulation in the slow stress rate could assist infants' tracking of speech by increasing rhythmic cues. Natural speech is not perfectly regular. This lack of clear rhythm is a challenge for the synchronization between neural activity and speech input. In adults, linguistic knowledge can compensate for the lack of rhythm by top-down modulating auditory activity via linguistic predictions (Keitel et al., 2017; Rimmele et al., 2018; Meyer et al., 2019; Ten Oever and Martin, 2021). Yet, preverbal infants still lack the linguistic knowledge required for such predictions. The enhancement of slow amplitude modulations in IDS could compensate for this lack by providing additional acoustic cues which aids tracking for the prosodic stress rate. A second possibility is that IDS modulates tracking by increasing infants' attention, possibly via a combination of visual and acoustic cues. The typical acoustic correlates of IDS were shown to increase infants' attention compared to ADS (Consortium, 2020; Kaplan et al., 1995; Roberts et al., 2013; Cooper and Aslin, 1990). Neural tracking is affected by attention (Fuglsang et al., 2017) and reflects the selection of relevant attended information (Obleser and Kayser, 2019). Increased tracking of IDS in the prosodic stress rate may thus reflect 9-month-olds' enhanced attention to prosodic stress, which provides them with a relevant acoustic cue aiding word segmentation. These two interpretations are not mutually exclusive but may explain our findings as a combination of enhanced acoustic cues in maternal speech and increased attention of the infant for prosodic stress in IDS.

One question that we cannot account for is whether the enhanced synchronization between neural activity and IDS observed here results from genuine entrainment of endogenous oscillations or from auditory-evoked reponses (Keitel et al., 2021, see). It has been suggested that oscillations in the auditory cortex phase-lock to acoustic information in a frequency specific manner (Lakatos et al., 2013). In speech processing, F0 amplitude rhythms might entrain neural oscillations in the delta frequency (Bourguignon et al., 2013). For our current results, this could indicate that the amplitude edges or peaks in the prosodic stress rate of IDS provide sufficient rhythmic cues to allow for a phase-alignment of oscillatory activity operating in the frequency range of prosodic stress. Another possibility is that the exaggeration of prosodic stress in IDS leads to a series of evoked responses that are superimposed on neural activity and thus appear in the same frequency band as the prosodic stress rate. Our results are compatible with both explanations, therefore future work is required to distinguish these two accounts for infants' processing of IDS. Since both possbilities result in increased neural processing of acoustic information in the prosodic stress rate in IDS, they are also both compatible with our interpretation that tracking facilitates infants' word segmentation from continuous IDS.

Our study provides further evidence for the previously proposed importance of prosody in assisting speech processing. This is especially relevant in light of healthy parent-infant interactions given evidence that clinically depressed mothers show less IDS, potentially impacting children's language development (Lam-Cassettari and Kohlhoff, 2020; Stein et al., 2008; Liu et al., 2017). In healthy parent-infant interactions, IDS may be optimally adapted to infants' needs during language development (see Kalashnikova and Burnham, 2018). As infants grow older, the amount of parents' IDS decreases and changes its acoustic characteristics (Kitamura and Burnham, 2003; Raneri et al., 2020). Leong et al. (2017) showed that the enhancement of prosodic amplitude modulations in IDS decreases when mothers are talking to older infants. These changes in IDS may be tied to infants' increased linguistic knowledge, as parents were shown to use more prototypically infant-directed speech when talking to infants with lower language abilities (Reissland and Stephenson, 1999; Kalashnikova et al., 2020; Bohannon and Marquis, 1977). Importantly, speech tracking was shown to increase with linguistic knowledge (Chen et al., 2020; Choi et al., 2020), meaning that infants' tracking may rely less on acoustic cues in IDS as their linguistic knowledge increases. This implies that parents adapt the acoustic properties of their speech to their infants' language development to allow for a level of tracking that is optimal for the infants' current language status. Future studies need to evaluate the interactions between parents' speech adaptations and infants' linguistic knowledge on infants' tracking of speech. The current study contributes an empirical foundation for such future investigations, by showing that neural tracking is sensitive to parents' speech adaptations during natural interactions, likely facilitating higher-level inferential processes such as word segmentation. This makes tracking a potential neural mechanism for infants' word segmentation from continuous speech.

## Declaration of Competing Interest

The authors declare that there is no conflict of interest.

## Funding

This work was supported by the Max Planck Society. The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

## Acknowledgements

We are grateful to the infants and parents who participated.

## References

Adriaans, F. and Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America*, 141(5):3070–3078.

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., et al. (2021). Delta-and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, page 118698.

Begus, K., Southgate, V., and Gliga, T. (2015). Neural mechanisms of infant learning: differences in frontal theta activity during object exploration modulate subsequent object recognition. *Biology letters*, 11(5):20150041.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9):341–345.

Bohannon, John Neil, I. and Marquis, A. L. (1977). Children's control of adult speech. *Child Development*, pages 1002–1008.

Bourguignon, M., Baart, M., Kapnoula, E. C., and Molinaro, N. (2020). Lip-reading enables the brain to synthesize auditory features of unknown silent speech. *Journal of Neuroscience*, 40(5):1053–1065.

Bourguignon, M., De Tiege, X., De Beeck, M. O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., and Jousmäki, V. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, 34(2):314–326.

Castellanos, N. P. and Makarov, V. A. (2006). Recovering eeg brain signals: Artifact suppression with wavelet enhanced independent component analysis. *Journal of Neuroscience Methods*, 158(2):300–312.

Çetinçelik, M., Rowland, C. F., and Snijders, T. M. (2020). Do the eyes have it? a systematic review on the role of eye gaze in infant language development. *Frontiers in Psychology*.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, 5(7).

Chen, Y., Jin, P., and Ding, N. (2020). The influence of linguistic information on cortical tracking of words. *Neuropsychologia*, 148:107640.

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., and Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science*, 31(9):1161–1173.

Consortium, M. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, 3(1):24–52.

Cooper, R. P. and Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, 61(5):1584–1595.

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Linguistics and Language Compass*, 7(3):157–170.

Crosse, M. J., Butler, J. S., and Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *Journal of Neuroscience*, 35(42):14195–14204.

Csibra, G. and Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. *Processes of change in brain and cognitive development. Attention and performance XXI*, 21:249–274.

Csibra, G. and Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, 13(4):148–153.

Cutler, A. and Carter, D. M. (1987). The predominance of strong initial syllables in the english vocabulary. *Computer Speech & Language*, 2(3-4):133–142.

de Cheveigné, A. (2020). Zapline: A simple and effective method to remove power line artifacts. *NeuroImage*, 207:116356.

Delorme, A. and Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85:761–768.

Fernald, A. and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20(1):104.

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

Fuglsang, S. A., Dau, T., and Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage*, 156:435–444.

Gabard-Durnam, L. J., Mendez Leal, A. S., Wilkinson, C. L., and Levin, A. R. (2018). The harvard automated processing pipeline for electroencephalography (happe): Standardized processing software for developmental and high-artifact data. *Frontiers in Neuroscience*, 12:97.

Goswami, U. (2019). Speech rhythm and language acquisition: An amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences*, 1453:67–78.

Graf Estes, K. and Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, 18(5):797–824.

Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., and Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*, 53(6):1529–1542.

Grieser, D. L. and Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, 24(1):14.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., and Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, 11(12).

Háden, G. P., Mády, K., Török, M., and Winkler, I. (2020). Newborn infants differently process adult directed and infant directed speech. *International Journal of Psychophysiology*, 147:107–112.

Hoehl, S., Michel, C., Reid, V. M., Parise, E., and Striano, T. (2014). Eye contact during live social interaction modulates infantsâĂŹ oscillatory brain activity. *Social Neuroscience*, 9(3):300–308.

Jessen, S., Fiedler, L., Münte, T. F., and Obleser, J. (2019). Quantifying the individual auditory and visual brain response in 7-month-old infants watching a brief cartoon movie. *NeuroImage*, 202:116060.

Jessen, S., Obleser, J., and Tune, S. (2021). Neural tracking in infants–an analytical tool for multisensory social processing in development. *Developmental Cognitive Neuroscience*, page 101034.

Junge, C., Cutler, A., and Hagoort, P. (2014). Successful word recognition by 10-month-olds given continuous speech both at initial exposure and test. *Infancy*, 19(2):179–193.

Junge, C., Kooijman, V., Hagoort, P., and Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science*, 15(4):463–473.

Jusczyk, P. W., Houston, D. M., and Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cognitive Psychology*, 39(3-4):159–207.

Kalashnikova, M. and Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of Child Language*, 45(5):1035âĂŞ1053.

Kalashnikova, M., Goswami, U., and Burnham, D. (2020). Infant-directed speech to infants at risk for dyslexia: A novel cross-dyad design. *Infancy*, 25(3):286–303.

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., and Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports*, 8(1):1–8.

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R., Owren, M. J., and Cooper, R. P. (1995). Dishabituation of visual attention by infant-versus adult-directed speech: Effects of frequency modulation and spectral composition. *Infant Behavior and Development*, 18(2):209–223.

Katz, G. S., Cohn, J. F., and Moore, C. A. (1996). A combination of vocal f0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development*, 67(1):205–217.

Keitel, A., Ince, R. A., Gross, J., and Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *NeuroImage*, 147:32–42.

Keitel, C., Obleser, J., Jessen, S., and Henry, M. J. (2021). Frequency-specific effects in infant electroencephalograms do not require entrained neural oscillations: A commentary on köster et al.(2019). *Psychological Science*, page 09567976211001317.

Kitamura, C. and Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, 4(1):85–110.

Kooijman, V., Hagoort, P., and Cutler, A. (2009). Prosodic structure in early word segmentation: Erp evidence from dutch ten-month-olds. *Infancy*, 14(6):591–612.

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., and Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology*, 4:25.

Lakatos, P., Musacchia, G., OâĂŹConnel, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, 77(4):750–761.

Lam-Cassettari, C. and Kohlhoff, J. (2020). Effect of maternal depression on infant-directed speech to prelinguistic infants: Implications for language development. *Plos one*, 15(7):e0236787.

Leong, V. and Goswami, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PLoS ONE*, 10(12):1–37.

Leong, V., Kalashnikova, M., Burnham, D., and Goswami, U. (2017). The temporal modulation mtructure of Infant-Directed Speech. *Open Mind*, 1(2):78–90.

Liu, Y., Kaaya, S., Chai, J., McCoy, D., Surkan, P., Black, M., Sutter-Dallay, A.-L., Verdoux, H., and Smith-Fawzi, M. (2017). Maternal depressive symptoms and early childhood cognitive development: a meta-analysis. *Psychological medicine*, 47(4):680–689.

Lloyd-Fox, S., Széplaki-Köllőd, B., Yin, J., and Csibra, G. (2015). Are you talking to me? neural activations in 6-month-old infants in response to being addressed during natural interactions. *cortex*, 70:35–48.

Männel, C. and Friederici, A. D. (2013). Accentuate or repeat? brain signatures of developmental periods in infant word recognition. *Cortex*, 49(10):2788–2798.

Martin, A., Igarashi, Y., Jincho, N., and Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition*, 156:52–59.

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, 48(7):2609–2621.

Meyer, L., Sun, Y., and Martin, A. E. (2019). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, 0(0):1–11.

Michel, C., Matthes, D., and Hoehl, S. (2021). Neural and behavioral correlates of ostensive cues in naturalistic mother-infant interactions. Manuscript in preparation.

Obleser, J. and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, 23(11):913–926.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: Open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011.

Ortiz Barajas, M. C., Guevara, R., and Gervain, J. (2021). The origins and development of speech envelope tracking during the first months of life. *Developmental Cognitive Neuroscience*, 48:100915.

Parks, T. and McClellan, J. (1972). Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Transactions on Circuit Theory*, 19(2):189–194.

Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6):1378–1387.

Power, A. J., Mead, N., Barnes, L., and Goswami, U. (2012). Neural entrainment to rhythmically presented auditory, visual, and audio-visual speech in children. *Frontiers in Psychology*, 3:216.

RamÃŋrez-Esparza, N., GarcÃŋa-Sierra, A., and Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, 17(6):880–891.

Raneri, D., Von Holzen, K., Newman, R., and Bernstein Ratner, N. (2020). Change in maternal speech rate to preverbal infants over the first two years of life. *Journal of Child Language*, 47(6):1263–1275.

Reissland, N. and Stephenson, T. (1999). Turn-taking in early vocal interaction: a comparison of premature and term infantsâĂŹ vocal interaction with their mothers. *Child: Care, Health and Development*, 25(6):447–456.

Rimmele, J. M., Morillon, B., Poeppel, D., and Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, 22(10):870–882.

Roberts, S., Fyfield, R., Baibazarova, E., van Goozen, S., Culling, J. F., and Hay, D. F. (2013). Parental speech at 6 months predicts joint attention at 12 months. *Infancy*, 18:E1–E15.

Schreiner, M. S. and Mani, N. (2017). Listen up! Developmental differences in the impact of IDS on speech segmentation. *Cognition*, 160:98–102.

Singh, L., Nestor, S., Parikh, C., and Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, 14(6):654–666.

Smith, N. A. and Strader, H. L. (2014). Infant-directed visual prosody: MothersâĂŹ head movements and speech acoustics. *Interaction studies*, 15(1):38–54.

Smith, S. M. and Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1):83–98.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416(6876):87–90.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27(4):501–532.

Soderstrom, M., Blossom, M., Foygel, R., and Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35(4):869–902.

Spinelli, M., Fasolo, M., and Mesman, J. (2017). Does prosody make the difference? a meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, 44:1–18.

Stärk, K., Kidd, E., and Frost, R. L. (2021). Word segmentation cues in German child-directed speech: A corpus analysis. *Language and Speech*, (February).

Stein, A., Malmberg, L.-E., Sylva, K., Barnes, J., Leach, P., and team, F. (2008). The influence of maternal depression, caregiving, and socioeconomic status in the post-natal year on children's language development. *Child: care, health and development*, 34(5):603–612.

Swerts, M. and Krahmer, E. (2010). Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics*, 38(2):197–206.

Ten Oever, S. and Martin, A. E. (2021). An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions. *eLife*, 10:e68066.

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1):53–71.

Weisleder, A. and Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11):2143–2152.

Winkler, I., Haufe, S., and Tangermann, M. (2011). Automatic classification of artifactual ica-components for artifact removal in eeg signals. *Behavioral and Brain Functions*, 7(1):1–15.

Zangl, R. and Mills, D. L. (2007). Increased brain activity to infant-directed speech in 6-and 13-month-old infants. *Infancy*, 11(1):31–62.

redit Auth

**Katharina Menn**: Conceptualization - current study, Formal Analysis, Visualization, Writing - Original Draft. **Christine Michel**: Conceptualization - initial study, Investigation, Data Curation, Writing - Review & Editing. **Lars Meyer**: Conceptualization - current study, Formal analysis, Writing - Original Draft, Supervision. **Stefanie Hoehl**: Conceptualization - initial study, Resources, Writing - Review & Editing. **Claudia Männel:** Conceptualization - current study, Supervision, Writing - Original Draft.

Data availability

The conditions of our ethics approval do not permit public archiving of participant data. Readers seeking access to the data should contact the lead author Katharina Menn to arrange a formal data sharing agreement.

Code availability

Preprocessing of the EEG data was done using the publicly available HAPPE pipeline V1 (DOI: 10.3389/fnins.2018.00097; download: https://github.com/lcnhappe/happe) in EEGLAB v2019.1 (DOI: https://doi.org/10.1515/bmt-2013-4182; download: https://sccn.ucsd.edu/eeglab/download.php) and in fieldtrip (version from 20200521) (DOI: https://doi.org/10.1155/2011/156869; download: https://www.fieldtriptoolbox.org/download.php). Custom code was written for the computation of speech envelopes and Hilbert coherence and will be made available if the article is accepted for publication.