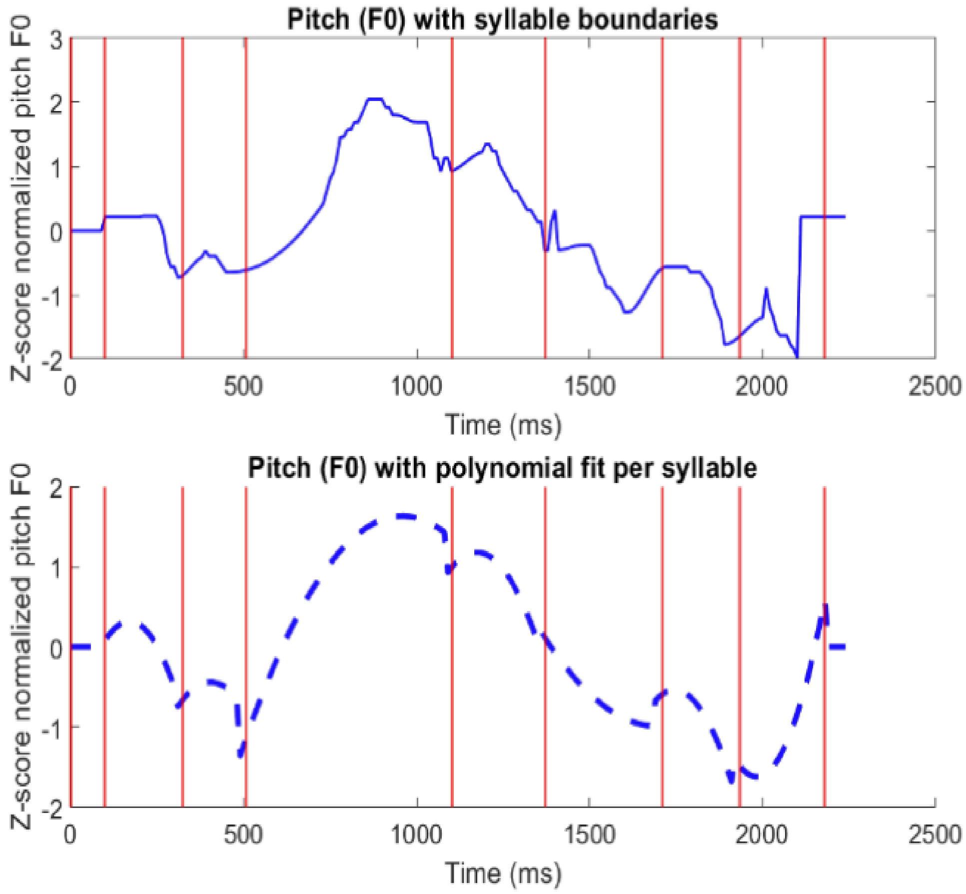
Does Pitch Surprisal Drive Infants' Attention to Infant-directed Speech and Song?

Jule Hafermann
Supervisors: Dr. Marlene Meyer and Dr. Tineke Snijders
Second Reader: Dr. Eelke Spaak

Radboud University, The Netherlands

July 2024


Pitch (F0) with syllable boundaries
Z-score normalized pitch F0
3
2
1
0
-1
-2
0
500
1000
1500
2000
2500
Time (ms)
Pitch (F0) with polynomial fit per syllable
Z-score normalized pitch F0
2
1
0
-1
-2
0
500
1000
1500
2000
2500
Time (ms)

Abstract

Infant-directed speech (IDS) and infant-directed songs (ID songs) are two important ways caregivers interact with their infants. IDS and ID songs have also been found to attract infants' attention and promote language learning. This study investigates the proposal that IDS and ID songs capture infants' attention by containing higher amounts of pitch surprisal than adult-directed speech. To test this idea, we analyze pitch surprisal throughout IDS and ID song stimuli and assess whether surprisal modulates 10-month-old infants' frontal theta power, a neural marker of attention. Even though we find high levels of surprisal in IDS and ID song, our study does not support that surprisal drives frontal theta power. Alternatively, we suggest that pitch surprisal might be tracked at higher levels, like words, and should be considered in interaction with its linguistic content. While our study provides some evidence that theta power reflects overall attentional engagement, the mechanism underlying its fluctuations is not understood yet.

1 Introduction

1.1 Does Pitch Surprisal Drive Infants' Attention to Infant-directed Speech and Song?

When speaking to infants, adults naturally adapt their speech to what is called infant-directed speech (IDS). In contrast to IDS, adult-directed speech (ADS) describes the speech used when addressing another adult. Even when they are unfamiliar with the language spoken, adult participants across the world can identify whether speech is infant-directed or not (Mehr et al., 2019). Therefore, we can conclude that IDS shares at least some acoustic features across languages and that these features are different from the acoustic properties of ADS. A growing body of evidence supports that infant-directed verbal engagement is crucial as a source of input for developing language skills (e.g. Thiessen et al., 2005; Liu et al., 2003, Singh et al., 2009; Ma et al., 2011). This direct verbal engagement has mainly included IDS but may extend to infant-directed songs (hereafter ID song), which most infants are exposed to daily (Yan et al., 2021;Hahn et al., 2028). Why IDS and ID songs are more beneficial for language development than ADS is still debated. Here, we focus on the recent proposal that the acoustic property of the perceived height of sounds (pitch) is more variable in IDS and ADS and thereby, contains higher degrees of surprisal (Räsänen et al., 2018). Combining computational modelling with neuroimaging measures, we test whether higher surprisal of pitch in IDS and ID song could promote speech processing through greater attentional engagement.

1.2 The Role of IDS in Language Development

From adult ratings and acoustic analyses, we know that IDS shares features across languages that differ from ADS (Mehr et al., 2019; Hilton et al., 2022). These features include a higher overall pitch, more exaggerated pitch contouring, the use of short phrases, many pauses, and questions (e.g. Soderstrom et al., 2008). Importantly, infants can pick up on these differences. Evidence for this comes from the multi-lab ManyBabies study during which infants from across the globe listened to English IDS and ADS stimuli (ManyBabies Consortium, 2020). They found that infants reliably preferred to attend to the IDS sounds. These findings have been consistent across different ages of infants and experimental set-ups (see Zettersten et al., 2024 for a recent review). Going even further, IDS is suggested to be important for language development. Positive impacts of caregivers using IDS when interacting with their infant are observable in various aspects of language acquisition, including word segmentation (e.g. Thiessen et al., 2005), speech discrimination (e.g., Liu et al., 2003), phoneme categorization (e.g., Kuhl et al., 1997; Werker et al., 2007), and word learning (e.g. Singh et al., 2009; Ma et al., 2011).

1.3 ID Song as a Second Source of Language Input

Next to IDS, ID song is very common in infant-caregiver interactions (Ilari, 2005; Yan et al., 2021). For instance, Hahn et al. (2018) report that 93% of Dutch infants are exposed to children's songs daily.

Songs can vary in their functions and are often divided into play songs for entertainment and lullabies that serve to soothe the infant (Cirelli et al., 2019). However, whether they are meant to soothe or entertain, all ID songs also provide verbal input for the infant. Combined with their frequent use, ID songs offer a second important source of direct verbal input for the infant.

Comparing IDS to ID songs, we observe similar effects on infants' cognition. First, ID songs capture infants' attention as well as IDS (Corbeil et al., 2013; Costa-Giomi and Ilari, 2014). Some studies even report preferences for ID songs over IDS (e.g. Tsang et al., 2016). Second, ID songs can also have positive effects on language learning. For instance, parental reports of high levels of singing with their 6-month-old infants predicted higher language skills of the infant during their second year of life (Franco et al., 2021). Children's speaking and listening skills also benefitted from music interactions through a 20-week parent-child music program, as reported by parents (Harris, 2011). In addition, musical skills, including for example rhythm perception and production, were linked to better language skills in preschoolers (Politimou et al., 2019).

To summarize, previous research found that infants preferentially attend to IDS and ID songs and that both have benefits for various aspects of language learning. While language learning benefits might not seem directly related to attention, there is evidence that the IDS benefits stem from greater attention to the speech stream rather than direct facilitation of the processing of linguistic units (Zhou et al., 2023). In line with this, the dynamic attention framework suggests that IDS promotes attention to speech through neural entrainment (Nencheva and Lew-Williams, 2022). Neural entrainment describes the synchronization of oscillations in the brain with an external stimulus, which in our study is IDS. This is thought to direct infants' attention to relevant moments in the sound stream. Nencheva and Williams (2022) propose that the combination of a more predictable rhythm in IDS together with a more variable pitch leads to stronger entrainment in IDS than ADS. Strong entrainment to IDS, as well as increased salience through pitch variations, could then underlie infants' sustained attention to this type of stimulus. Since rhythm in songs is even more prominent than in speech, we can hypothesize that the mechanism of neural entrainment might underlie attention to ID songs as well.

1.4 Surprise! – A Mechanism for Infants' Attention Preferences for IDS and ID Songs

As described in the dynamic attention framework (Nencheva and Lew-Williams, 2022), it has recently been suggested that IDS better captures infants' attention by having a more surprising pitch than ADS (Räsänen et al., 2018). The notion of surprisal comes from Shannon's Information Theory (Shannon, 1948), where surprisal is defined as the negative log probability of an event within its context. It is measured in shannons where one shannon is equal to the probability of a coin landing on heads ($p = 0.5$). Here, we view surprisal within the predictive processing framework. In this framework, our brain is described as a prediction machine (Köster et al., 2020). By constantly making predictions about the expected sensory input (e.g. what word we will heard next when listening to someone speak), the brain aims to reduce the amount of uncertainty in our dynamic environment. This process can be divided into two steps. First, we generate predictions of sensory input from an internal model of our environment and compare this to the actual input. Second, we update our model depending on how closely our predictions matched the actual input. The larger the difference between our predictions and the actual input, the more we have to update our model. Situating surprisal in this framework, a more surprising stimulus is a stimulus which creates a larger difference between our predictions and the actual input. More surprising stimuli seem ideal for learning since they create the strongest need to update our internal model. In line with this, Räsänen et al. (2018) suggest that IDS contains higher amounts of surprisal than ADS, making it better suited for learning. Recognizing IDS as a greater learning opportunity might then explain why infants preferentially attend to IDS when comparing it to ADS.

As proposed by Räsänen et al. (2018) and the dynamic attention framework (Nencheva and Lew-Williams, 2022), one aspect of IDS that might be more surprising than ADS is its pitch contouring. Pitch itself is defined as the perceived height of a sound (Editors of Encyclopaedia Britannica, 2023). Pitch contouring then describes how the height of the voice moves up or down over time. More broadly, pitch contouring is part of the prosody of language, which refers to properties like stress, rhythm, and intonation, that determine the sound of speech beyond the meaning of single words (American Psychological Association, 2014). Infants are sensitive to prosodic features from birth. For instance, newborns are already able to recognize their mother's voice (DeCasper and Fifer, 1980), as well as the rhythm of their native language (Mehler et al., 1988). Therefore, it is plausible that infants pick up on differences in the variability of pitch contouring in IDS and ADS. Applying computational modelling to a large corpus of IDS and ADS speech recordings, Räsänen et al. (2018) support that pitch contours are more surprising in IDS than ADS.

While higher pitch surprisal has been suggested to be a driving force of infants' attention in the theory of dynamic attention, as well as in the computational modelling approach by Räsänen et al. (2018), we are lacking empirical studies that test the proposed theories. Our study aimed to address this gap by testing whether infants are indeed sensitive to these differences in pitch surprisal and whether surprisal drives infants' attention to IDS and ID song.

1.5 Infant Frontal Theta Power as a Neural Marker of Attention

In infants, EEG frontal theta oscillations have been suggested as a neural indicator of attentional processes (Begus and Bonawitz, 2020). Theta oscillations typically span the 4-8 Hz range in adults and are slower in infants with a range of approximately 3-6 Hz (Orekhova et al., 1999; Saby and Marshall, 2012). Frontal theta power modulations have been found in response to internally controlled allocation of attention (Orekhova et al., 1999). In a more naturalistic setting of free play, frontal theta power predicted the duration of visual fixations (Wass et al., 2018). There is also evidence that frontal theta power reflects sustained attentional engagement (Xie et al., 2018; Braithwaite et al., 2020; Jones et al., 2020) In speech processing, increases in infants' theta power have been found during the processing of syllables in a foreign language (Bosseler et al., 2013). This was thought to reflect attentional engagement, as well as cognitive effort associated with processing new information. Therefore, frontal theta power seems to be suitable as a measure of attentional engagement. However, it should be noted that theta power has been used as a neural marker in infant research in other domains as well, including the surprisal of a stimulus itself, learning and memory, social learning, and exploration (Begus and Bonawitz, 2020).

While the relation between frontal theta power and surprisal in IDS and ID songs has not been investigated yet, a recent study has found that for the degree of surprisal of infant-directed actions modulated theta power (4-5Hz) over frontal and central electrodes (Meyer et al., 2022). Thus, infants seem to be sensitive to surprisal in the visual domain, leading us to expect similar patterns in other infant-directed interactions.

1.6 The Current Study

Aiming to expand our understanding of the neural underpinnings of infant-directed interactions and what drives the attentional preference and learning benefits of IDS and songs, our study asks whether frontal infant theta power (3-6 Hz) fluctuates as a function of the moment-by-moment degree of surprisal of pitch in IDS and ID songs. We specifically decided to focus on the surprisal of the pitch contours of syllables. Syllables consist of a vowel with or without additional consonants surrounding it (Britannica, 2024). They have been indicated as important structural units in early language processing (e.g. Bijeljac-Babic et al., 1993; Bertoncini and Mehler, 1981; Gomez et al., 2014). Infants seem to already be able to track speech at the syllable rate (Menn et al., 2022). In addition, there is evidence for tracking of IDS and ID song at the theta oscillation rate which is similar to the syllable rate (Menn et al., 2022; Leong et al., 2017; Attaheri et al., 2022). Thus, we hypothesized that surprisal variations at the syllable level could drive modulations in the ongoing theta oscillations. We expected to find a positive relationship between infant theta power and pitch surprisal showing that first, infants are sensitive to pitch surprisal, and second, surprisal increases infants' attention to IDS and ID song.

To test our hypothesis, we re-analysed data from an infant EEG study by Snijders et al. (2020). During this study, infants listened to IDS and ID songs. First, we extracted the average frontal theta power per syllable from the EEG data. Second, we used an adapted version of the Mixed-Order Markov Chain (MOMC) model by Räsänen et al. (2018) to calculate the surprisal per syllable for the IDS and ID song stimuli. We then ran a linear mixed effects model (LMM) to test whether theta power fluctuates as a function of the degree of surprisal of pitch.

2 Methods

2.1 Extracting Average Theta Power per Syllable

2.1.1 Participants

The participants were 40 Dutch 10-month-old infants. They were all born at 37-42 weeks gestational age and had no known neurological or language impairments in the immediate family. All participants were recruited via a database of families who volunteer in research at the Baby and Child Research Center, Nijmegen, the Netherlands. Informed consent was obtained from the caregivers. Each infant then completed two experimental sessions. After data collection, one infant was excluded from the analysis due to being raised bilingual. In addition, individual sessions were excluded for having too many bad channels (n = 5) or too few trials (<20) remaining after initial preprocessing (n = 14). The final number of infants included in the analysis was 31 and the total number of experimental sessions was 59.

2.1.2 Materials

The EEG experiment was divided into two sessions: one for IDS and one for ID songs. Within a session, infants were presented with 20 familiarization stimuli in either ID song or IDS. A familiarization stimulus consisted of eight phrases. To assess the learning of specific target words from the familiarization stimuli, there were also one-phrase-long spoken test stimuli. The test stimuli were presented as groups of four stimuli after each familiarization stimulus. In total, there were 80 possible familiarization stimuli (40 ID songs/40 IDS) and 80 test stimuli. For every ID song familiarization stimulus, there was an identical IDS stimulus. In addition, the same eight phrases, except for the target words, were used for another ID song and IDS stimulus (for an example, see Table 1). The melodies of the ID song stimuli came from German, English, French, Norwegian, and Dutch children's songs which were reported to be unknown by 22 native Dutch caregivers of 10-month-old infants. The original lyrics were replaced by the Dutch familiarization phrases. Figure 1 shows an example of an ID song stimulus. All stimuli were recorded by a trained female singer in a sound-attenuated booth using the audio software Adobe Audition. Later editing of the audio stimuli, including annotation of words and syllables, was done using the speech analysis tool Praat (Bohaček and Cepanec, 2023, version 6.4.12). The mean phrase length of the stimuli was 5.71 words (range 3-10) and 7.82 syllables (range 4-14). The full materials are available in the supplementary information of Snijders et al. (2020).

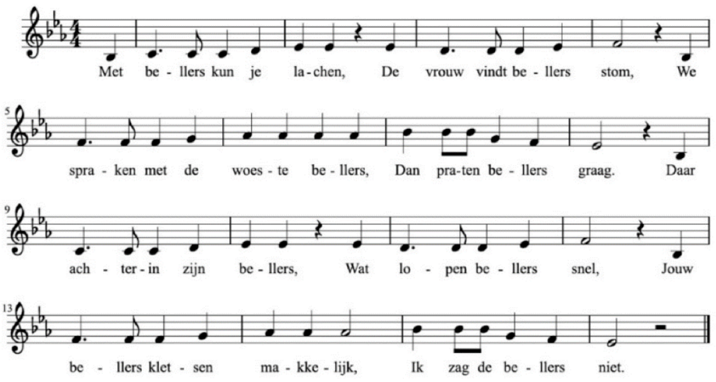


Figure 1: Example Melody and Text of ID Song Stimulus

Note: Example of song familiarization stimulus with melody and lyrics (in Dutch) written underneath. The melodies were taken from children's songs and the text consisted of eight phrases per song. Figure from Snijders et al. (Snijders et al., 2020).

Table 1: Example of Familiarization and Test Stimuli in IDS and ID Song Session

Session *(first or second)*	**Familiarization Phase** *(in IDS or ID Song)*	**Test Phase** *(always in IDS)*
IDS	Met bellers kun je lachen.	Aan die piefen gaf hij koffie.
	De vrouw vindt bellers stom.	Vaak gaan bellers op reis.
	We spraken met de woeste bellers.	Alle bellers stappen laat uit.
	Dan praten bellers graag.	Zij zijn goede piefen geworden.
	Daar achterin zijn bellers.	
	Wat lopen bellers snel!	
	Jouw bellers kletsen makkelijk.	
	Ik zag de bellers niet.	
ID Song	Met piefen kun je lachen.	Vaak gaan bellers op reis.
	De vrouw vindt piefen stom.	Zij zijn goede piefen geworden.
	We spraken met de woeste piefen.	Alle bellers stappen laat uit.
	Dan praten piefen graag.	Aan die piefen gaf hij koffie.
	Daar achterin zijn piefen.	
	Wat lopen piefen snel!	
	Jouw piefen kletsen makkelijk.	
	Ik zag de piefen niet.	

Note: Every infant completed an IDS and an ID song session. In the IDS session, familiarization stimuli were presented in IDS. In the ID song session, they were presented as ID song. Each familiarization stimulus was followed by four spoken test stimuli in the test phase to assess the learning of a target word (here bellers in blue for IDS, piefen in green for ID song) from the familiarization phase. Two test phrases contained the target word from that session, while two contained the target word that would be presented during familiarization in the other session. The familiarization phrases of the two sessions differed in the target words (here piefen vs bellers) and in the modality (IDS vs ID song) in which they were presented.

2.1.3 Procedure

The infants completed the song and speech session of the experiment during two lab visits which were separated by an average of 7.6 days (range 5-14 days). Each session took around one hour with the experiment lasting around 20 minutes. The order of the sessions was counterbalanced across infants. After the infant got used to the lab environment and the caregiver received information about the experiment, the infant was fitted with a pre-gelled EEG cap. Extra gel was added during impedance checking if necessary. The experiment was then conducted in a sound-attenuated booth with a Faraday cage. The infant sat on their caregivers' lap while the stimuli were presented over two loudspeakers at 65 dB. There was also a screen showing an image which was not linked to the auditory stimuli, as well as an experimenter operating toys to keep the infant engaged. The caregiver and experimenter operating the toys listened to unrelated music over headphones. A second experimenter outside the booth was responsible for the EEG data acquisition and stopping the experiment when the infant became upset. The stimuli were presented using the Presentation® software (Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). The familiarization stimuli were divided into eight experimental versions of 20 trials. The order of the versions was counterbalanced across participants. Each trial consisted of a familiarization phase which was immediately followed by the corresponding test phase (for an example, see Table 1). Before each familiarization and test phase, an auditory attention cue in the form of the phrase 'Luister eens!' (English: Listen to this!) was presented to the infant. For familiarization, this phrase was spoken in the speech session and sung in the song session to ease the transition between modalities. The phrase was always spoken before the test phase of a trial.

2.1.4 EEG Recordings

EEG recordings were conducted using 32 active Ag/AgCl electrodes (ActiCAP, 10-20 system), Brain Amp DC, and Brain Vision Recorder software (Brain Products GmbH, Germany). The electrode used as the online reference was FCz. Electro-oculogram (EOG) recordings were made using Fp1 (above eye), and F9 and F10 (outer sides of eyes). The recorded electrodes were F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP6, TP10, P7, P3, Pz, P4, P8, PO9, and Oz. The data was sampled at 500 Hz and filtered online with a time constant of 10 s and a high cutoff at 1000 Hz. Usually, electrode impedances did not exceed 25 kΩ.

2.1.5 EEG Preprocessing

The EEG data analysis was performed using MATLAB (The MathWorks, Natick, MA, USA) and the MATLAB software toolbox ‘Fieldtrip’ (Oostenveld et al., 2011). Upon request, all scripts for the preprocessing and frontal theta power analysis are available on
GitHub at https://github.com/JuleHH/Jule$_{T}hesis_{I}DS_{I}DSong.$

The EEG data were filtered from 0.1 to 30 Hz and segmented into 1-second trials for initial preprocessing. Trials with large artifacts were rejected via visual inspection and on an individual basis. Next, eye movement and noise components were identified using an independent component analysis (ICA). The data were then re-preprocessed with trials of 3 seconds with a pre-stimulus period of 0.5 seconds and a post-stimulus period of 2.5 seconds. Trials were only created for stimuli that the infant finished listening to. The first trial of a stimulus was defined as the onset of the speech, song, or test stimulus. Additional trials were then created by shifting the onset of the trial to the onset of the next syllable. This was repeated until the last syllable of the stimulus. This resulted in overlapping trials of 3 seconds, with one trial per syllable. The trial length was chosen based on the longest relevant data segment which was a pause of 2.18 seconds. The identified ICA components were then removed and the data was re-referenced to the linked mastoids. Segments with additional artifacts above and below 150uV were removed using visual artifact rejection, and flat channels were excluded. Lastly, some of the channels were repaired (on average 0.7 channels per dataset, range 0 to 3).

2.1.6 Frontal Theta Power Analysis

For analysing fluctuations in frontal theta power, we identified Fz, FC1, FC2, FCz, and Cz as our electrodes of interest to calculate mean power values (see Figure 2). In a follow-up analysis of Snijders et al. (2020), frontal theta power was found at these electrodes of interest. Therefore, we chose the same electrodes for our analysis. Our frequency range of interest was 3-6 Hz which has been shown to correspond to the theta frequency band in infants (Orekhova et al., 1999; Saby and Marshall, 2012). We then conducted a time-frequency analysis with a Hanning taper, window length one second, and a sliding time window of 100 ms. The window length allowed for an estimation of theta power at a 1Hz frequency resolution. We first estimated theta power at 3, 4, 5, and 6Hz for every 100ms after trial onset up to 600ms. To extract the mean frontal theta power per syllable, we calculated the average theta power for our electrodes and frequencies of interest between 200-400ms and 400-600 ms after syllable onset. From previous literature, it is not very clear when we could expect changes in theta power in reaction to individual syllables during the presentation of continuous speech or song. Coming closest to our design, we used Bosseler and colleagues’ (2013) study as an indication who found increases in infant theta power in response to syllables in the time window of 200-600ms.

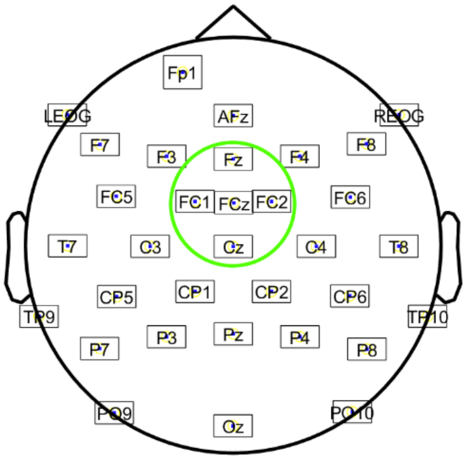


Figure 2: Electrode locations and electrodes of interest

Note: Topographic plot of the locations of the electrodes (F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP6, TP10, P7, P3, Pz, P4, P8, PO9, Oz, LEOG, REOG, Fp1) during EEG data collection. The green circle shows the electrodes of interest, namely Fz, FC1, FC2, FCz, and Cz.

2.2 Modelling of Surprisal in IDS and ID Song

To obtain surprisal values for the individual syllables, we adapted the computational model by Räsänen et al. (2018) who investigated the predictability P of syllables in IDS and ADS. In preparation of our auditory stimuli for the computational model, we took the following steps (see Figure 3): (1) manual syllabification of stimuli using Praat (Boersma and Weenink, 2024, version 6.4.12), (2) estimating pitch trajectories of stimuli, (3) dividing stimuli into syllable segments, (4) parametrizing the pitch trajectories for each syllable segment, (5) apply unsupervised k-means clustering to syllable parameters to categorize pitch shapes. We then used a Mixed-order Markov chain (MOMC) model to estimate the predictability of the pitch of each syllable in its context. Lastly, we converted this estimate of P to surprisal S using the following formula based on Shannon's Information Theory (Shannon, 1948): $S_{syllable} = -log_2(P_{syllableincontext})$.

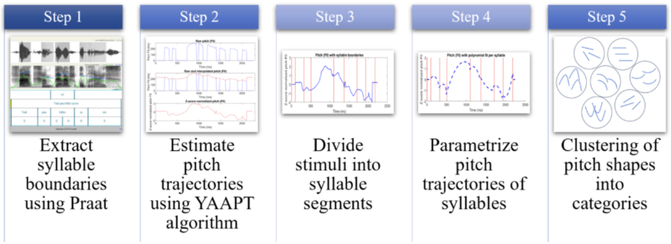


Figure 3: Steps to Extract Pitch Trajectory Parameters of Stimuli

2.2.1 Extracting Parameters of Pitch (F0) Trajectories

Due to the model's syllable algorithm leading to noticeable deviations from the original syllable boundaries, we first decided to manually annotate the syllables of our song and speech stimuli in Praat (Boersma and Weenink, 2024, version 6.4.12). A syllable boundary here refers to the start or end time of a syllable which we specified in Praat. Subsequently, the boundaries were saved in a text grid and further processed in Matlab using the mPraat toolbox (Bořil and Skarnitzl, 2016).

As a second step, we applied the YAAPT pitch algorithm (Zahorian Hu, 2008; version 4.0) to extract the fundamental frequency (F0) trajectories per stimulus (see Figure 2). Pitch was estimated at a 100 Hz sampling rate in the range of 120-600Hz. In contrast to the Räsänen et al. (2018) stimuli, our stimuli contained longer pauses between each of the 8 phrases. To avoid assigning pitch values to these pauses, we only applied the YAAPT pitch fix tool within each phrase. The pitch fix checked the estimated pitch trajectories for estimation errors and then interpolated the pitch trajectories across unvoiced parts of the phrase. To allow for a comparison with the various speakers of the stimuli from Räsänen et al. (2018), the F0 trajectories were transformed into a logarithmic scale and z-score normalized ($M = 0, SD = 1$).

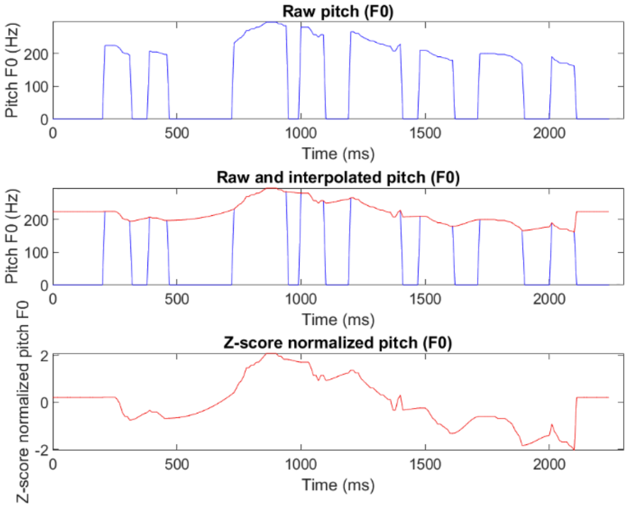


Figure 4: Raw Pitch (F0) Interpolation and Normalization

Note: Top: The raw pitch values for an example stimulus extracted using the pitch algorithm YAAPT. Middle: The raw pitch values in blue and the interpolated pitch values in red. Bottom: We z-score normalized the pitch values (M= 0, SD =1) for comparability with the Räsänen et al. (2018) study, which used z-score values to account for variations in the overall voice height of the different speakers.

In the third step, the stimuli were segmented into syllables according to the syllable boundaries extracted from Praat (Boersma and Weenink, 2024, version 6.4.12) (Figure 3).

Fourth, the F0 trajectories were parametrized for each syllable using a second-order polynomial fit. Taking out the constant, the polynomial coefficients were taken as parametric descriptions of the shapes of F0 syllable segments.

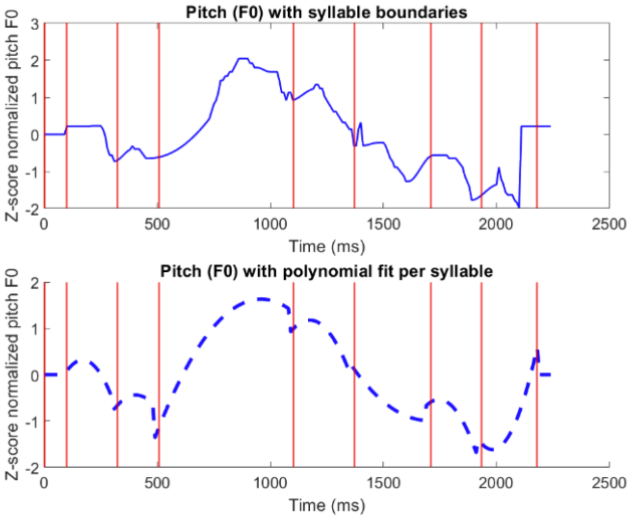


Figure 5: Segmentation and Parametrization of Stimuli

Note: Top: Example of interpolated pitch in blue with syllable boundaries indicated in red. Bottom: We extracted the shape parameters of the pitch contour (blue) for each syllable segment (boundaries in red) using a second-order polynomial fit (blue dashed line).

Fifth, using k-means clustering with random initialization, the parameters were vector quantized into 6, 12, or 24 categories (Q). The three different values were chosen to reduce the impact of choosing a specific number of categories and the predictability results of the three different models were later averaged. The outcome of the preprocessing was a description of the F0 trajectory of a stimulus as a sequence of the various F0 shapes now categorized as states q with syllable segment s being described as qs.

2.2.2 Extracting Predictability and Surprisal Values

We used a Mixed-order Markov chain (MOMC) model to estimate the predictability of each syllable. In our MOMC, the predictability P of an F0 shape qs in the syllable segment s is given by

$$P(q_s|q_{s-1}, ..., q_{s-m}) = \Sigma_{k=1}^{m} {}_k(q_{s-k}) M_k(q_{s-k}, q_s) \Pi_{k-1}^{j=1}[1 -_j (q_{s-j})]$$

where M_k is a k-lag-specific transition matrix with transition weight k. M describes the transition probabilities between the segments at different lags k and weighs them according to their reliability in the syllable segments context. The order of the MOMC was $m = 5$ based on Räsänen et al. (2018) meaning that the pitch shape of syllable segment s is predicted based on the preceding five syllable pitch shapes. The stimuli were then split into 90% training and 10% testing segments employing random sampling with 50% of training samples from the song stimuli, and 50% from the speech stimuli. After the training of the model, the predictability was estimated for the remaining testing segments. The training-testing process was repeated in a 10-fold manner until all stimuli had been used in the testing sample. The outcome of the MOMC model was a predictability estimate ($P_{syllableincontext}$) of each syllable segment. As a final step, the predictability values P were transformed into surprisal values which were calculated as $S_{syllable} = -log_2(P_{syllableincontext})$.

2.3 Statistical Analysis

The statistical analysis was conducted using the statistical programming language R (R Core Team, 2024) in the RStudio environment (Posit Team, 2024). Upon request, all analysis scripts are available on GitHub at https://github.com/JuleHH/Jule_Thesis_IDS_IDSong.

In the first analysis, we compared the average surprisal of pitch contours from our stimuli (IDS and ID song) to the average surprisal of the stimuli used in Räsänen et al. (2018). We first conducted a one-way ANOVA with surprisal as the dependent variable and stimulus type (ADS vs IDS from Räsänen et al., 2018 vs IDS from Snijders et al., 2020 vs ID song) as the independent variable. Due to unequal variances and violation of homoscedasticity, we repeated the analysis using a Kruskal-Wallis test.

To investigate the relationship between frontal theta power (3-6Hz) and surprisal and its interaction with stimulus type (song vs speech), we compared two linear mixed-effects models (LMMs) using the lme4 package (Bates et al., 2015). Significance tests were conducted using the lmerTest package (Kuznetsova et al., 2017). The models were fitted using restricted maximum likelihood (REML) and t-tests employed Satterthwaite's method for degrees of freedom.

The difference between the two models was the time of interest of theta power which was 200-400ms for Model 1 and 400-600ms for Model 2. In both models, theta power was the outcome variable which was predicted from surprisal S. In addition, we included an interaction of surprisal with stimulus type (IDS vs ID song) to test for differences in theta and the surprisal-theta relationship between

IDS and ID song. We also included trial number as a fixed effect since some studies have found that theta power increases over longer periods of stimulus presentations, for instance when infants were watching one-minute-long dynamic videos (Jones et al., 2020; Editors of Encyclopaedia Britannica, 2023). Individual infants were included as a random effect. To account for potential differences in baseline theta power in the two different sessions, session number (1 or 2) was included as a random effect nested within the individual infant. In building up the model, random slopes per individual subjects were included but removed for the final model since their random effect variance estimates were nearly zero and led to convergence issues. Due to the log-normal distribution of raw theta power, we applied a logarithmic transformation to theta power. We then z-score standardized theta power and the continuous predictors of surprisal and trial number to avoid model convergence issues resulting from the outcome variable and the predictors being on very different scales. The final two model equations looked as follows:

Model 1:

$$ThetaPower(200\text{–}400ms) = \beta0 + \beta1(Surprisal) + \beta2(StimulusType) +$$

$$\beta3(Surprisal \times StimulusType) + \beta4(TrialNumber) + u_{Infant} + u_{SessionwithinInfant} + \varepsilon$$

Model 2:

$$ThetaPower(400\text{–}600ms) = \beta0 + \beta1(Surprisal) + \beta2(StimulusType) +$$

$$\beta3(Surprisal \times StimulusType) + \beta4(TrialNumber) + u_{Infant} + u_{SessionwithinInfant} + \varepsilon$$

where β0 is the intercept, β1-β4 the coefficients of the different predictors, u_{Infant}, the random intercept for each infant, $u_{SessionwithinInfant}$, the random intercept of every session nested within each infant, and ε the error term of the model.

3 Results

3.1 Surprisal in IDS and ID Song

In total, there were 160 stimuli (40 songs, 40 IDS, 80 test phrases) with a total number of 6335 syllables, 3527 for speech and 2808 for song. The average surprisal of syllables in the song stimuli was 2.15 ($SD = 1.20$) with values ranging from 0.35 to 8.43. The average surprisal of syllables in the speech conditions (IDS, test) was 2.86 ($SD = 1.26$) with values ranging from 0.35 to 8.95. We then compared the surprisal of the IDS and ID song stimuli to the average surprisal of IDS and ADS syllables from Räsänen et al. (2018) (see Table 2). The Kruskal-Wallis test with surprisal as the dependent variable and stimulus type (ADS vs IDS from Räsänen et al. (2018) vs IDS from Snijders et al. (2020) vs ID Song) as the independent variable showed that there was a significant difference between the stimulus types (${}^2(3) = 3355.6, p < .001, {}^2 = .097$). A post-hoc pairwise comparison using Dunn's test with Bonferroni correction revealed that there was a significant difference between all stimulus types (all $p < .001$). The average surprisal was highest in the IDS stimuli from Snijders et al. (2020) and lowest in the ADS stimuli. ID songs were higher in surprisal than ADS and IDS stimuli from the Räsänen et al. (2018) study but lower than the IDS stimuli from Snijders et al. (2020) (see Figure 6).

Table 2: Descriptives of Surprisal of Pitch Contours in ADS, IDS, and ID Song

	Räsänen et al. (2018)		Snijders et al. (2020)	
	ADS	IDS	IDS	ID Song
n	16535	11816	3527	2808
mean (sd)	1.64 (1.33)	2.09 (1.53)	2.86 (1.26)	2.15 (1.20)
min	0.13	0.12	0.35	0.35
max	11.79	11.08	8.95	8.43
IQR	1.35	1.89	1.56	1.49

Note: The table shows the descriptives of the calculated surprisal values from the stimuli used in the Räsänen et al. (2018) and Snijders et al. (2020) studies. Surprisal is measured in shannons. N indicates the number of syllables for which surprisal was calculated. IQR = interquartile range.

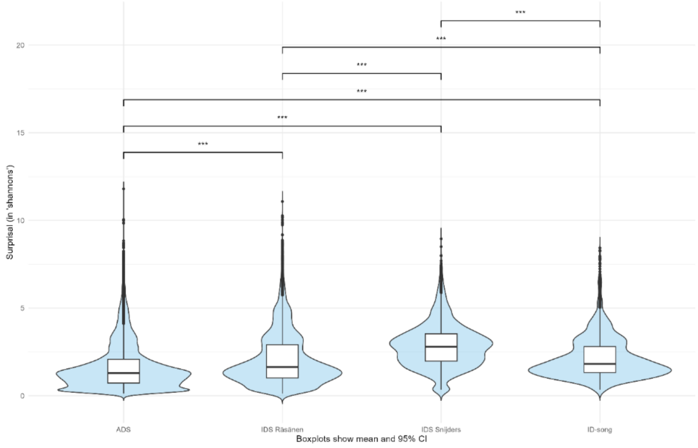


Figure 6: Distribution of Surprisal Values in ADS, IDS, and ID Song

Note: The figure shows the distribution of surprisal (in shannons) from the stimuli used in the Räsänen et al. (2018) and Snijders et al. (2020) studies. The boxplots indicate the mean and 95% confidence intervals. A Kruskall-Wallis test with post-hoc pairwise comparisons revealed that all means differ significantly from each other (p < .001). On average, ADS was lowest on surprisal, followed by the IDS stimuli from Räsänen et al. (2018) and then the ID songs. The average surprisal was highest for the IDS stimuli from Snijders et al. (2020).

3.2 The Effect of Surprisal on Frontal Theta Power

To evaluate the effect of surprisal on frontal theta power, we compared 2 linear mixed effects models. The first model predicted theta power with a lag of 200-400ms after syllable onset from surprisal S of the syllable, stimulus type (IDS vs ID song), and trial number. Individual infant and session variations were taken into account by allowing for random intercepts per session (first or second) nested in infants. The second model included the same fixed and random effects but predicted the theta power with a lag of 400-600ms after syllable onset.

A total of 71222 observations were included in the two LMMs from 59 sessions and 31 individual subjects. For a detailed overview of the observations per condition and subject. The average surprisal and distribution of surprisal values from the included trials for both stimulus types was similar to the initial analysis of surprisal in IDS and ID song (see Table 3 and Figure 7).

Table 3: Descriptives of Surprisal in IDS and ID Song for LMMs

	IDS	ID Song
n	49591	21631
mean (sd)	2.92 (1.36)	2.14 (1.19)
min	0.35	0.35
max	8.95	8.43
IQR	1.66	1.49

Note: A total of 71222 trials from 31 infants were included in the LMMs. A trial consisted of one syllable. The table shows the number of trials per stimulus type (IDS and ID song) and the mean with standard deviation, as well as the smallest (min), largest (max), and interquartile range (IQR) of the surprisal values of the trials included. Surprisal is measured in shannons.

Figure 7: Distribution of Surprisal Values in IDS and ID Song for LMMs

Note: The figure shows the distribution of surprisal values (in shannons) of IDS and ID song which were included in the two LMMs. The distribution was similar to the distribution from the MOMC model meaning that the values for the LMMs were a good representation of the MOMC results (for comparison, see Figure 6).

The LMMs revealed similar results (for summary, see Table 4). Pitch contour surprisal did not predict theta power at a lag of 200-400ms ($= -0.002, SE = 0.004, t(71170) = -0.492, p = .623$) (Figure 8), nor at a lag of 400-600ms ($= -0.006, SE = 0.004, t(71170) = -1.541, p = .123$). For both models, there was no significant effect of the condition (IDS or ID song) on theta power, nor an interaction between the condition and surprisal. There was a significant effect of trial number on theta power at a lag of 200-400ms ($= 0.035, SE = 0.004, t(70610) = 9.457, p < .001$) (Figure 9) and at a lag of 400-600ms ($= 0.037, SE = 0.004, t(70590) = 9.75, p < .001$). Overall, the fixed effects explained only 0.1% of the variance in the data ($marginal R^2 = .001$), while the combined fixed and random effects explained around 19% of the variance ($conditional R^2 = .189$ for Model 1 and $R^2 = .19$ for Model 2). Thus, higher pitch surprisal of a syllable did not predict higher frontal theta power 200-400ms or 400-600ms post syllable onset. Furthermore, whether a syllable was sung or spoken did not predict changes in theta power. It also did not influence the relationship between surprisal and theta power. Finally, the results showed that theta power increased over time, as indicated by trial number.

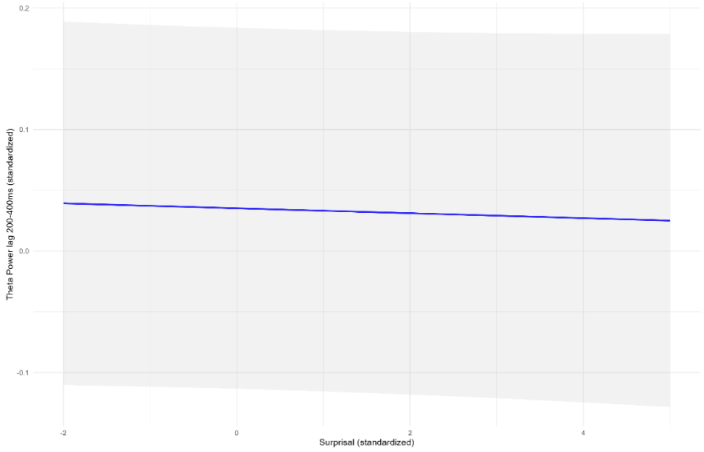


Figure 8: Model Predictions for Relationship between Frontal Theta Power (200-400ms lag) and Surprisal

Note: The figure shows the model-based estimates for the relationship between standardized theta power at a lag of 200-400ms and standardized surprisal. 95% CI are indicated in grey. The figure illustrates that there is no relationship between surprisal and theta power. The estimates were calculated using the ggeffects package in R (Lüdecke, 2018).

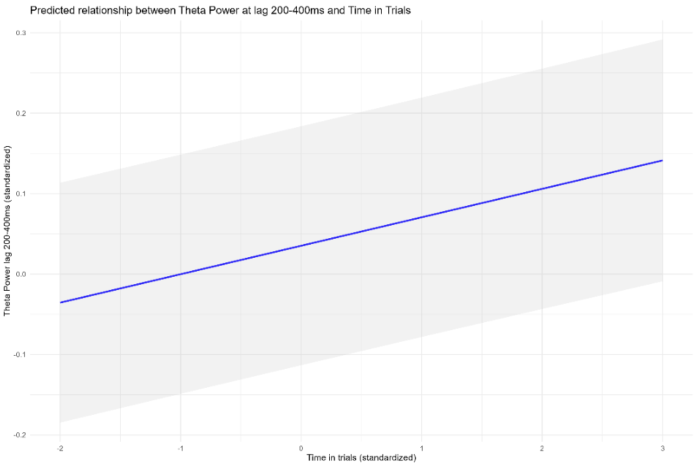


Figure 9: Model Predictions for Effect of Trial Number on Frontal Theta Power (200-400ms lag)

Note: The figure shows the model-based estimates for the relationship between standardized theta power at a lag of 200-400ms and standardized time in trials. 95% CI is indicated in grey. The figure shows that there is an increase in frontal theta power as time increases. This effect was significant. The estimates were calculated using the ggeffects package in R (Lüdecke, 2018).

Table 4: Summary of LMM Parameter Estimates for Predictors and Random Effects

	Model 1: Theta Power 200–400ms					**Model 2: Theta Power 400–600ms**				
Predictors	**Estimate**	**SE**	**CI**	**df**	**t-value**	**Estimate**	**SE**	**CI**	**df**	**t-value**
(Intercept)	0.04	0.08	-0.11–0.18	29	0.47	0.04	0.08	-0.11–0.18	29	0.48
Surprisal (standardized)	-0.002	0.004	-0.01–0.01	7117	-0.49	-0.006	0.004	-0.01–0.00	7117	-1.54
Condition (IDS vs ID song)	0.01	0.01	-0.01–0.03	3903	0.92	0.008	0.01	-0.01–0.03	3858	0.68
Trial number	0.035	0.004	0.03–0.04	7061	9.46***	0.037	0.004	0.03–0.04	7959	9.75***
Surprisal × Condition	0.0065	0.008	-0.01–0.02	7116	0.79	0.009	0.008	-0.01–0.02	7116	1.07
Random Effects	**Variance**		**Std. Dev.**			**Variance**		**Std. Dev.**		
Individual subject	0.16		0.39			0.15		0.39		
Session nested within subject	0.04		0.20			0.04		0.20		
Residual	0.84		0.92			0.84		0.92		

$^{***}p < .001$

3.3 Visual Exploration of Individual Differences

Our model indicated that most of the variation in frontal theta power could be attributed to individual differences between the infants. To get an impression of these individual differences, we visually inspected the relationship between frontal theta power and surprisal per individual infant and session. Especially for song, there seems to be greater variation in the individual relationships between surprisal and theta power (Figure 10). For the effect of trial number, we also see that not all of the infants showed a positive relationship between trial number and frontal theta power (Figure 11) This was the case for IDS and ID song and at both time lags.

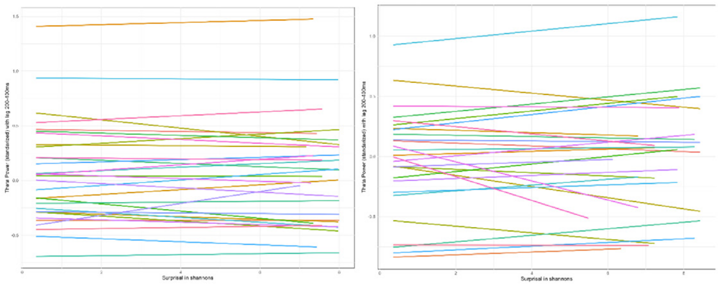


Figure 10: Individual Differences in the Relationship Between Surprisal and Theta Power

Note: The figure shows the relationship between standardized frontal theta power and surprisal in shannons for IDS (left) and ID song (right). Each colour indicates an individual infant.

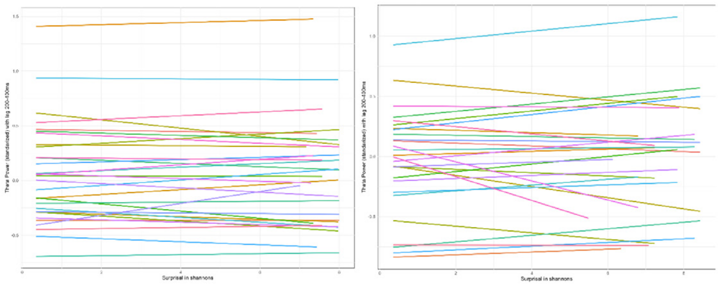


Figure 11: Individual differences in the Relationship between Trial Number and Theta Power

Note: The figure shows the relationship between standardized frontal theta power and trial number for IDS (left) and ID song (right). Each colour indicates an individual infant.

4 Discussion

Infants preferentially attend to IDS and ID songs when contrasted with ADS (e.g. ManyBabies Consortium, 2020; Corbeil et al., 2013; see Zettersten et al., 2024 for a recent review on IDS). However, what drives infants' attention to IDS and ID songs is not known yet. Here, we investigated whether greater pitch surprisal in IDS than ADS can explain greater attention to IDS. We tested this idea by first, calculating the amount of pitch surprisal for IDS and ID song syllables and by second, relating surprisal to infants' frontal theta power (3-6Hz) as a neural indicator of attention.

4.1 Pitch Surprisal in IDS, ID Song, and ADS

In the first part of the analysis, we compared the average surprisal of pitch contours of syllables in ID songs and IDS. We found that surprisal was significantly higher in IDS than ID songs. Both averages were higher than the ADS and IDS surprisal values from Räsänen and colleagues (2018), who used the same model to extract surprisal values for syllables from the ManyBabies dataset (ManyBabies

Consortium, 2020). The greater amounts of pitch surprisal in IDS and ID songs than in ADS are in line with the idea that infants' attention might be driven by the degree of surprisal of a stimulus (Räsänen et al., 2018; Meyer et al., 2022). It also fits with the dynamic attention account, which suggests that infants' preferential attention to IDS and ID song could be due to an interaction of temporal regularity with more variability in pitch (Nencheva and Lew-Williams, 2022).

While surprisal was higher for both, ID songs and IDS, than ADS, the difference between IDS and ID songs was also significant. IDS pitch contours showed a higher degree of surprisal than the ones of ID song. Finding this difference is interesting since behavioural studies have shown similar attention to IDS and ID song in infants (Corbeil et al., 2013; Costa-Giomi and Ilari, 2014) or even greater attentional preferences for ID song (Tsang et al., 2016). However, if surprisal drives attention, our results would predict that infants preferentially attend to IDS rather than ID songs. In contrast with the predictions from our model, but in line with previous research, we did not find differences between frontal theta power as a marker of attention for IDS vs ID song.

A possible explanation for higher surprisal in IDS than ID song is the coupling of pitches to syllables in songs. In our stimuli, a syllable typically corresponded to one musical note (for an example, see Figure 1). A musical note has a fixed pitch and pitch changes usually occur from one note to the next. As a consequence, the sung syllable corresponding to a musical note will also have a rather stable pitch. Since we did not take jumps in average pitch between syllables into account, there might have been fewer variations in pitch shapes for the ID song than the IDS stimuli in our study. This could have decreased the average surprisal of the songs. In our study, we chose to focus on pitch movements within a syllable to be able to compare the results to the results by Räsänen et al. (2018). To find out whether syllables in IDS and ID song differ in overall pitch surprisal, a comparison model could be run predicting average pitch height in addition to the pitch shape.

A second finding of our MOMC model was that our stimuli were higher in surprisal than the ADS and IDS stimuli included in the Räsänen et al. (2018) model. One reason for this difference could be that we had less data available, which led to higher surprisal due to a smaller amount of training data given to the model. A second reason could be that Räsänen et al. (2018) used pitch interpolation over the whole length of the stimuli including pauses between different phrases. In our study, we only used interpolation within each phrase and assigned zero values to the pauses. Predicting not only pitches but also the absence of pitch could contribute to the overall higher surprisal. Making this choice, our study might more accurately reflect how infants make predictions in continuous speech or song, which usually consist of multiple phrases separated by pauses. Pauses and phrase length are two aspects of speech that have been found to differentiate IDS from ADS, with IDS containing shorter phrases and longer pauses than ADS (Cox et al., 2023). Furthermore, several studies report age-related changes in pauses and phrases in IDS (Bohaček and Cepanec, 2023). Pauses tend to become shorter with increasing age of the infant and phrase length changes with development as well. Therefore, investigating how infants at different ages predict not only pitch but also the absence of pitch, can contribute to a more complete understanding of speech and song processing in early development.

To summarize, our MOMC model of surprisal in a different set of stimuli confirmed that IDS pitch contours contain higher levels of surprisal than ADS. In addition, we expanded these findings to ID songs, which show higher surprisal than ADS but lower surprisal than IDS. Overall, we found a higher average surprisal in IDS than Räsänen et al. (2018) which calls for further research into the variations of surprisal in different sets of stimuli and the factors driving these differences.

4.2 Pitch Surprisal and Frontal Theta Power

Our main research question was whether pitch surprisal drives infants' attention, for which we used frontal theta power (3-6Hz) as a neural indicator. Specifically, we investigated whether the surprisal of a syllable pitch contour predicted theta power at a lag of 200-400ms and 400-600ms. We hypothesized that higher surprisal would predict higher frontal theta power. However, we did not find such an effect of surprisal on theta power. In addition, frontal theta power and the relationship between surprisal and theta power did not differ depending on the type of stimulus, which was either IDS or ID song. Our results show a significant positive effect of the trial number on theta power at both times of interest. This means that theta power tended to increase the longer infants listened to the stimuli.

Contrary to our expectations, our study does not support the idea that pitch surprisal drives attention to IDS or ID songs, as indicated by fluctuations in frontal theta power. First, this could mean that the degree of pitch contour surprisal of each syllable does not determine infants' attention to speech. Instead, attention could be driven by predictions at the word level or might depend on age-related preferences. Second, it could mean frontal theta power as a neural marker does not reflect moment-to-moment attention at the syllable level.

4.2.1 The Role of Pitch in Attention

Not finding an effect of the degree of pitch contour surprisal on infants' attention raises the question of how important syllable-level pitch contour variations are in capturing infants' attention. Variability in pitch has been found to be characteristic of IDS (Cox et al., 2023; Fernald et al., 1989; Fernald Simon, 1984; Genovese et al., 2019; Grieser Kuhl, 1988; Narayan McDermott, 2016; Stern et al., 1983) and pitch specifically seems to be a crucial factor in driving infants' attention to IDS (Fernald and Kuhl, 1987; Segal and Newman, 2015). This makes it unlikely that pitch itself is not important in infants' processing of IDS and ID song.

One explanation for not finding this effect could be that pitch contours are not predicted at the syllable level. As an alternative, pitch might be predicted at the word level. For instance, Nencheva et al. (2021) found that pitch variability at the word level is also greater in IDS than in ADS, and that pitch contours tend to fall into four distinct categories (rise, fall, hill, and valley). Thus, it could be the case that our model, which used syllables and 6, 12, and 24 categories of pitch shapes did not accurately reflect pitch predictions in natural language processing.

A second reason could be that the IDS and ID song stimuli were not a good representation of caregivers' input to infants at the age of 10 months. Snijders et al. (2020) reported that the IDS stimuli were rather fast compared to stimuli in other IDS studies. In addition, caregivers have been found to adapt their IDS to the age of the infant (Bohaček and Cepanec, 2023). Next to this, infants' preferences for certain aspects of IDS seem to change with development. Kitamura and Notley (2009) found that 6-month-olds preferred words containing hill-shaped vowel pitch contours and stretched vowels, which are typical for IDS, while 10-month-olds preferred monotone vowel contours and non-stretched vowels. Therefore, not finding an effect of pitch surprisal on attention could also be due to the changing nature of IDS throughout development.

4.2.2 The When and How of Frontal Theta Power

Next to additional potential sources of surprisal and their interaction with pitch, it is also not clear whether our measure of frontal theta power accurately reflected attentional engagement at the syllable level. While frontal theta power fluctuations have been found in relation to surprisal in infant-directed actions (Meyer et al., 2022) and attentional engagement with visual and auditory stimuli (Wass et al., 2018; Bosseler et al., 2013), there is a lot of variability in the use of theta power as a neural marker (Begus and Bonawitz, 2020).

One issue pertains to the temporal dynamics of frontal theta power. Several studies have found an increase in theta power prior to stimulus onset (Orekhova et al., 1999) and that this increase predicted the length of the following visual fixation (Wass et al., 2018). Other studies have reported fluctuations in frontal theta power after the onset of the stimulus (Meyer et al., 2022; Bosseler et al., 2013). However, a majority of these studies employed designs with longer inter-stimulus intervals which allow for easier linking of a specific stimulus to the associated frontal theta power. For studies that used continuous stimuli, theta power is often calculated over longer intervals and not related to single events (e.g. Jones et al., 2020; Braithwaite et al., 2020). Coming closest to our design, we used Bosseler and colleagues' (2013) study as an indicator of changes in theta power in relation to the auditory processing of syllables. We also investigated two separate time windows of interest (200-400ms and 400-600ms post-stimulus onset) since we did not have strong expectations for when we would see the associated changes in frontal theta power. Our results of not finding an effect of surprisal on frontal theta power between 200 and 600ms after syllable onset could indicate that the effect occurs outside of the 200-600ms window. To better inform the use of frontal theta power as a neural marker of attentional engagement, further research into the mechanism and temporal dynamics of frontal theta power is needed.

While surprisal did not predict frontal theta power in our study, trial number did. Overall, infants' frontal theta power increased over the duration of an experimental session. This could indicate infants' attentional engagement with the stimuli. Infants' sustained attention to the stimuli fits with results from the original ERP analysis of this data that showed that infants were segmenting words from both, ID song and IDS during the familiarization phase (Snijders et al., 2020). In line with this, combinations of EEG with heart rate measures have shown that increases in frontal theta power can be associated with periods of sustained attention (Xie et al., 2017). Furthermore, several studies have reported links between increases in frontal theta power, learning, and later cognitive skills. For instance, Jones et al. (2020) found that frontal theta power in 12-month-old infants increased while watching dynamic videos and that this increase was video-specific, indicating learning. Furthermore, the extent of the increase was related to cognitive skills during childhood. These findings were confirmed by Braithwaite et al. (2020) who found a relationship between increased frontal theta power at 6 months and non-verbal cognitive skills at 9 months of age. Thus, there is evidence that our findings of increased frontal theta power can be associated with infants' engagement with the stimuli. However, from visual inspection of individual differences in the effect, we see that not every infant showed an increase in theta power over time. Knowing that increases in frontal theta power are related to attention and cognitive skills, it would be interesting to further explore whether these individual differences relate to differences in infants' sensitivity to variations in pitch surprisal in IDS and ID song and later language skills.

4.2.3 Limitations and Future Directions

In our study, we used an adapted MOMC model by Räsänen et al. (2018) to model the pitch contour surprisal of syllables in IDS and ID song. However, some model choices that were kept to ensure comparability between the different models might not have been appropriate for our set of stimuli. In songs, pitch might be rather stable within each syllable and a model predicting the average pitch per syllable could be more appropriate. Furthermore, our study solely focused on the order of the pitch shapes and disregarded temporal differences. For instance, syllables in the ID songs were typically longer than in IDS. In driving attention, the dynamic attention theory by Nencheva and Williams (2022) suggests that the slower, predictable rhythm of IDS interacts with more surprising pitch contours. Our study tested one aspect of this theory, namely whether pitch surprisal drives attention. However, the potential interaction with the temporal dynamics of speech and song remains to be explored.

Another limitation of the study was the more exploratory nature of the choice of time window for our frontal theta power analysis. We chose to investigate frontal theta power 200-600ms after syllable onset based on Bosseler et al. (2013). However, frontal theta power has been used as a marker of attention for times before stimulus onset to several seconds post-stimulus onset. It is important to address this variability and to come to a better understanding of frontal theta power in different study designs, e.g. single events compared to continuous stimuli. In addition, frontal theta power has also been linked to visual attention (e.g. Wass et al., 2018). In our study, an experimenter was in the room and operated toys to keep the infant engaged (if needed). Our data analysis did not control for moments where the infants' attention might have been drawn to the toys, experimenter, or caregiver. These additional sources for capturing infants' attention should be considered in future studies of IDS and ID song.

Finally, our visual exploration of the data revealed considerable individual differences. For example, while our model indicated an overall increase in frontal theta power over time, this was not the case for all infants. Some research has indicated that individual variations in theta power fluctuations are related to later cognitive skills (Jones et al., 2020; Braithwaite et al., 2020). In our study, there were no additional tests of the cognitive skills of the individual infants, for instance, receptive vocabulary size. Whether individual differences in attentional engagement as indicated by frontal theta power are related to language skills or even sensitivity to pitch surprisal could be explored in a future study.

5 Conclusion

In our study, we investigated whether the degree of surprisal of pitch contours in IDS and ID songs drives infants' attention, as measured by frontal theta power (3-6Hz). In line with this idea, we showed that IDS and ID song contain high degrees of pitch surprisal. We also found that frontal theta power increased the longer infants listened to the stimuli. This could indicate that infants were engaging with the stimuli. Our main finding was that pitch surprisal did not predict theta power 200-400ms and 400-600ms after the syllable onset. Thus, our study does not support that pitch surprisal drives infants' attention to IDS and ID song. The absence of this effect could be due to pitch not being predicted at the syllable level, as well as frontal theta power not reflecting moment-to-moment fluctuations in attention to a continuous sound stream. To conclude, our study contributes to the growing research into the mechanisms underlying attentional preferences for IDS and ID song. It also highlights the need for further research into the role of pitch in infants' speech processing and the mechanisms underlying frontal theta power as a neural marker for attention.

References

American Psychological Association. (2014). Prosody in speech and song. https://www.apa.org/pubs/highlights/peeps/issue-29

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., & Goswami, U. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, *247*, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698

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

Begus, K., & Bonawitz, E. (2020). The rhythm of learning: Theta oscillations as an index of active learning in infancy. *Developmental Cognitive Neuroscience*, *45*, 100810. https://doi.org/10.1016/j.dcn.2020.100810

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development*, *4*, 247–260.

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* (Version 6.4.15) [[Computer program]]. http://www.praat.org/

Bohaček, A., & Cepanec, M. (2023). ”it’s not just what we say or do, but how we say and do it”: A review of studies on infant-directed modifications. *Hrvatska Revija za Rehabilitacijska Istraživanja*, *59*(2), 81–98. https://doi.org/10.31299/hrri.59.2.5

Bořil, T., & Skarnitzl, R. (2016). Tools rpraat and mpraat. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, speech, and dialogue* (pp. 367–374). Springer International Publishing.

Bosseler, A. N., Taulu, S., Pihko, E., Mäkelä, J. P., Imada, T., Ahonen, A., & Kuhl, P. K. (2013). Theta brain rhythms index perceptual narrowing in infant speech perception. *Frontiers in Psychology*, *4*, 690. https://doi.org/10.3389/fpsyg.2013.00690

Braithwaite, E. K., Jones, E. J., Johnson, M. H., & Holmboe, K. (2020). Dynamic modulation of frontal theta power predicts cognitive ability in infancy. *Developmental Cognitive Neuroscience*, *45*, 100818. https://doi.org/10.1016/j.dcn.2020.100818

Corbeil, M., Trehub, S. E., & Peretz, I. (2013). Speech vs. singing: Infants choose happier sounds. *Frontiers in Psychology*, *4*, 372. https://doi.org/10.3389/fpsyg.2013.00372

Costa-Giomi, E., & Ilari, B. (2014). Infants’ preferential attention to sung and spoken stimuli. *Journal of Research in Music Education*, *62*(2), 188–194. https://doi.org/10.1177/0022429414530564

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2023). A systematic review and bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour*, *7*(1), 114–133. https://doi.org/10.1038/s41562-022-01452-1

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers’ voices. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

Editors of Encyclopaedia Britannica. (2023). Pitch [Encyclopedia Britannica]. https://www.britannica.com/topic/pitch-speech

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

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

Franco, F., Suttora, C., Spinelli, M., Kozar, I., & Fasolo, M. (2021). Singing to infants matters: Early singing interactions affect musical preferences and facilitate vocabulary building. *Journal of Child Language*, *49*(3), 552–577. https://doi.org/10.1017/s0305000921000167

Hahn, L. E., Benders, T., Snijders, T. M., & Fikkert, P. (2028). Infants' sensitivity to rhyme in songs. *Infant Behavior and Development*, *52*, 130–139. https://doi.org/10.1016/j.infbeh.2018.07.002

Harris, D. J. (2011). Shake, rattle and roll – can music be used by parents and practitioners to support communication, language and literacy within a pre-school setting? *Education 3-13*, *39*(2), 139–151. https://doi.org/10.1080/03004270903232691

Hilton, C., Moser, C., Bertolo, M., Lee-Rubin, H., Amir, D., Brainbridge, C., Simson, J., Knox, D., Glowacki, L., Alemu, E., Galbarczyk, A., Jasienska, G., Ross, C., Neff, M. B., Martin, A., Cirelli, L., Trehub, S., Song, J., Kim, M., & Mehr, S. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, *6*(11), 1545–1556. https://doi.org/10.1038/s41562-022-01410-x

Ilari, B. (2005). On musical parenting of young children: Musical beliefs and behaviors of mothers and infants. *Early Child Development and Care*, *175*(7-8), 647–660. https://doi.org/10.1080/0300443042000302573

Jones, E. J., Goodwin, A., Orekhova, E., Charman, T., Dawson, G., Webb, S. J., & Johnson, M. H. (2020). Infant eeg theta modulation predicts childhood intelligence. *Scientific Reports*, *10*(1), 11232. https://doi.org/10.1038/s41598-020-67687-y

Köster, M., Kayhan, E., Langeloh, M., & Hoehl, S. (2020). Making sense of the world: Infant learning from a predictive processing perspective. *Perspectives on Psychological Science*, *15*(3), 562–571. https://doi.org/10.1177/1745691619895071

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

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

Leong, V., Byrne, E., Clackson, K., Harte, N., Lam, S., De Barbaro, K., & Wass, S. (2017). Infants' neural oscillatory processing of theta-rate speech patterns exceeds adults' [Cold Spring Harbor Laboratory]. *bioRxiv*. https://doi.org/10.1101/108852

Liu, H., Kuhl, P. K., & Tsao, F. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*(3). https://doi.org/10.1111/1467-7687.00275

ManyBabies Consortium. (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. https://doi.org/10.1177/2515245919900809

Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, *251*, 118991.

Meyer, M., Van Schaik, J. E., Poli, F., & Hunnius, S. (2022). How infant-directed actions enhance infants' attention, learning, and exploration: Evidence from eeg and computational modeling. *Developmental Science*, *26*(1). https://doi.org/10.1111/desc.13259

Nencheva, M. L., & Lew-Williams, C. (2022). Understanding why infant-directed speech supports learning: A dynamic attention perspective. *Developmental Review*, *66*, 101047. https://doi.org/10.1016/j.dr.2022.101047

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). Fieldtrip: Open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Politimou, N., Bella, S. D., Farrugia, N., & Franco, F. (2019). Born to speak and sing: Musical predictors of language development in pre-schoolers. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00948

Posit Team. (2024). *Rstudio: Integrated development environment for r*. Boston, MA, Posit Software, PBC. http://www.posit.co/

R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna, Austria, R Foundation for Statistical Computing. https://www.R-project.org/

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising? – linking properties of ids to statistical learning and attention at the prosodic level. *Cognition*, *178*, 193–206. https://doi.org/10.1016/j.cognition.2018.05.015

Saby, J. N., & Marshall, P. J. (2012). The utility of eeg band power analysis in the study of infancy and early childhood. *Developmental Neuropsychology*, *37*(3), 253–273. https://doi.org/10.1080/87565641.2011.614663

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy*, *20*(3), 339–351. https://doi.org/10.1111/infa.12077

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, *14*(6), 654–666. https://doi.org/10.1080/15250000903263973

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

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

Tsang, C. D., Falk, S., & Hessel, A. (2016). Infants prefer infant-directed song over speech. *Child Development*, *88*(4), 1207–1215. https://doi.org/10.1111/cdev.12647

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in english and japanese. *Cognition*, *103*(1), 147–162. https://doi.org/10.1016/j.cognition.2006.03.006

Yan, R., Jessani, G., Spelke, E. S., De Villiers, P., De Villiers, J., & Mehr, S. A. (2021). Across demographics and recent history, most parents sing to their infants and toddlers daily. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1840). https://doi.org/10.1098/rstb.2021.0089

Zettersten, M., Cox, C., Bergmann, C., Tsui, A. S. M., Soderstrom, M., Mayor, J., Lundwall, R. A., Lewis, M., Kosie, J. E., Kartushina, N., Fusaroli, R., Frank, M. C., Byers-Heinlein, K., Black, A. K., & Mathur, M. B. (2024). Evidence for infant-directed speech preference is consistent across large-scale, multi-site replication and meta-analysis. *Open Mind*, *8*, 439–461. https://doi.org/10.1162/opmi_a_00134

Zhou, X., Wang, L., Hong, X., & Wong, P. C. M. (2023). Infant-directed speech facilitates word learning through attentional mechanisms: An fnirs study of toddlers. *Developmental Science*, *27*(1). https://doi.org/10.1111/desc.13424