

Infants' Lexical Processing: Independent Contributions of Attentional and Clarity Cues

Irena Lovcevic¹, Denis Burnham², Marina Kalashnikova^{3,4,2}

1. International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo Institutes for Advanced Studies, 7-3-1 Hongo Bunkyo-ku, Tokyo, 113-0033
Japan
2. The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith NSW 2751, Australia
3. BCBL. Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, San Sebastian-Donostia, Guipuzcoa, 2004, Spain
4. IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Author Note

We would like to thank the HEARing Cooperative Research Centre for the grant 82631, “The Seeds of Language Development”, to the 2nd author. The 1st author’s work was supported by World Premier International Research Center Initiative (WPI), MEXT, Japan. The 3rd author’s work was supported by the Basque Government through the BERC 2018-2021 program, by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S, and by the Spanish Ministry of Science and Innovation through the Ramon y Cajal Research Fellowship, RYC2018-024284-I. We thank the families for their valuable time.

Data and analyses reported in this paper are freely available at the following link:
https://osf.io/wb65q/?view_only=5c99b83d616d49c79c2ff5b89e388ed8.

Correspondence concerning this article should be addressed to: Irena Lovcevic, International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, 113-0033 Japan, irena.lovcevic@ircn.jp

Abstract

There is a long-standing debate in the literature about the benefits that acoustic components of Infant Directed Speech (IDS) might have for infants' language acquisition. One of the highly contested features is vowel space expansion, which refers to the enlargement of the acoustic space between the corner vowels /i, u, a/ in IDS compared to Adult Directed Speech (ADS). Some evidence indicates that vowel space expansion in IDS facilitates infants' speech perception, thus promoting language development, whereas other studies have questioned these benefits and have proposed that any processing benefits of IDS are due to its other prosodic features such as exaggerated and variable pitch. This study aimed to tease apart the effects of vowel space expansion and prosodic exaggeration in IDS on 18-month-old infants' speech processing. Using a looking-while-listening paradigm, two between-subjects conditions were compared: Exaggerated Pitch (with exaggerated pitch height and range, but without vowel space expansion) and Expanded Vowel Space (with vowel space expansion, but no exaggeration in pitch height and range). Our results showed that infants recognized the meanings of the words more accurately in the Expanded Vowel Space compared to the Exaggerated Pitch condition. This suggests that vowel space expansion in IDS facilitates infants' lexical processing even when it does not cooccur with the prosodic exaggeration typical of IDS.

Key words: infant-directed speech, lexical processing, vowel space expansion, prosodic exaggeration, looking-while-listening

When addressing infants, adults produce a specific speech register known as Infant Directed Speech (IDS) (Fernald & Simon, 1984). IDS has been classified as a form of *hyper speech* that yields several positive effects on infants' early emotional and cognitive development, and that promotes language acquisition (Fernald, 2000). One particular property of IDS that has been attributed this linguistic function is phonetic exaggeration. That is, given that caregivers exaggerate phonetic categories in IDS, it has been proposed that the resulting speech facilitates infants' speech processing (Kuhl et al., 1997). However, recent research has cast doubt on this claim, leading to the proposal that this phonetic exaggeration may not be developmentally significant since, rather than being produced by adults in response to the infant, it is instead a by-product of other IDS features such as prosodic modifications (McMurray et al., 2013), emotional speech or smiling (Benders, 2013), and vowel lengthening (Martin et al., 2015). To investigate whether phonetic exaggeration might be beneficial in infants' language development, this study directly tests the contribution of phonetic and prosodic exaggeration to facilitating lexical access and recognition of lexical forms in 18-month-old infants.

Measures of vowel space expansion are typically used to capture phonetic exaggeration in IDS (Kuhl, 2000). Vowel space expansion refers to the enlargement of the acoustic space between the three corner vowels /i, u, a/, and it is indexed by plotting these three vowels in two-dimensional Formant 1/Formant 2 (F1/F2) space and calculating the areas of the resulting triangles for IDS and Adult Directed Speech (ADS). There is extensive evidence that the vowel triangle area is significantly larger in IDS compared to ADS (e.g., Adriaans & Swingley, 2017; Burnham et al., 2002; Cristia & Seidl, 2014; Kalashnikova et al., 2017; Kuhl et al., 1997), and the consistency of this finding has been corroborated by two recent meta-analyses (Cox et al., 2022; Lovcevic et al., *in preparation*). This expansion of the acoustic space is proposed to improve speech intelligibility by providing easily

distinguishable acoustic differences between vowels (Bradlow et al., 1996). Hence, it is possible that speakers expand the vowel space in their speech in response to their audience's need for more intelligible speech. In fact, in support of this claim, vowel space expansion is not unique to IDS, but it is also found in other speech registers such as foreigner-directed speech (Lorge & Katsos, 2019; Piazza, et al., 2021; Uther, et al., 2007), read speech (Nakamura et al., 2008; Weirich & Simpson, 2019), and Lombard speech (Castellanos, et al., 1996; Hazan et al., 2015; Tang et al., 2017), hence, in speech where there might be greater need for more intelligible speech, but not in speech registers that solely serve an emotional function (e.g., pet-directed speech, Burnham et al., 2002, unless the pet is the parrot – a pet with some perceived linguistic potential, Xu et al., 2013).

In support of the linguistic benefits of vowel space expansion in infants' language acquisition, several studies have demonstrated that exposure to speech stimuli with expanded vowel spaces is positively related to vowel perception in six- to nine-month-old infants (Peter et al., 2016; Zhang et al., 2011), and to efficiency in spoken language processing in 19-month-old infants (Song et al., 2010). Furthermore, at the individual level, mothers' vowel space expansion has been linked to the development of their infants' speech perception skills (Kalashnikova & Carreiras, 2022; Liu et al., 2003), speech production abilities (Marklund et al., 2021), and vocabulary size (Hartman et al., 2017; Kalashnikova & Burnham, 2018; Lovcevic et al., 2020). The relation between potential benefits of vowel space expansion in IDS and maternal production of expanded vowel spaces might be explained via a social feedback loop (Warlaumont et al., 2014). Specifically, it is possible that maternal vowel space expansion in IDS is followed by speechlike vocalizations by an infant, which positively reinforces vowel space expansion in the mother's IDS and consequently further speech by the mother with expanded vowel space, leading to further speechlike vocalizations by an infant, and so on.

On the other hand, it has been argued that vowel space expansion may be a by-product of other acoustic features of IDS (prosody, smiling, and vowel lengthening, Benders et al., 2013; Martin et al., 2015; McMurray et al., 2013), so it does not serve any dedicated role in early language development. First, *a lack* of vowel space expansion in IDS has been documented for several languages including Dutch, Norwegian, Danish, and German (Audibert & Falk, 2018; Benders, 2013; Cox et al., 2022; Englund, 2018). Therefore, it appears that caregivers' tendency to expand vowel spaces when addressing infants is not universal. Nevertheless, it is noteworthy that patterns of phonetic exaggeration in this register may be dependent on the phonological inventories of each language, and it may be manifested in the exaggeration of other phonetic categories, as it has been shown for consonants (Englund & Behne, 2006) and lexical tones (Rattanasone et al., 2013). Second, even when vowel space expansion is present, there is greater dispersion and variability within vowel categories, which may actually result in *less intelligible* speech and complicate infants' task of learning the sound categories of their native language (Benders, 2013; Cristia & Seidl, 2014; Englund, 2018; McMurray et al., 2013). This high variability is attributed to the prosodic patterns in IDS, mainly slower speech rate and wider pitch range (McMurray et al., 2013). Supporting this claim, and contrary to studies showing processing advantages for vowels in IDS, evidence from computational studies shows more successful categorization of vowel sounds from ADS than from IDS input (Martin et al., 2015; McMurray et al., 2013; Miyazawa et al., 2017).

Prosodic exaggeration in maternal IDS is primarily manifested in increased mean height and range of fundamental frequency (F0) (Fernald & Simon, 1984; Han et al., 2020; Hilton et al., 2022; Trainor, et al., 2000; Wang et al., 2021) (note that this discussion is limited to the features of maternal IDS, which is predominantly described in the literature, while paternal IDS may have a different prosodic profile, Benders et al., 2021; Gergely et al.,

2017). This F0 feature of IDS has been shown to play a role in infants' emotional regulation (Fernald, 1993; Spinelli & Mesman, 2018; Stern et al., 1982), but it has also been proposed to aid language acquisition by attracting and maintaining infants' attention to the speech stream (Cooper & Aslin, 1990; Dunst et al., 2012; Fernald & Simon, 1984; ManyBabies Consortium, 2020). That is, heightened attention to speech may benefit language learning by increasing infants' arousal during exposure to speech input and priming their system for learning (Kaplan et al., 1996). Neurophysiological evidence supports this claim; IDS has been demonstrated to elicit higher levels of neural activity in newborns and nine-month-old infants using functional Near Infrared Spectroscopy (fNIRS) and electroencephalography (EEG) (Háden et al., 2020; Naoi et al., 2012; Saito et al. 2007; Santesso et al., 2007) as well as more efficient neural tracking of the speech signal in seven- and nine-month-olds using EEG (Kalashnikova et al., 2018; Menn et al., 2022). The use of stimuli with exaggerated prosody typical of IDS has also been shown to elicit more successful performance in a variety of language processing experimental tasks. Behavioral evidence demonstrates the facilitative effect of exaggerated pitch height and wider pitch range in IDS on infants' word segmentation (Thiessen et al., 2005), speech sound discrimination (Trainor & Desjardins, 2002), and novel word-referent mapping (Graf Estes & Hurley, 2013; Ma et al., 2011). Specifically, in relation to lexical processing, which is the main focus of this study, Zangl and Mills (2007) found, using EEG, that words produced with exaggerated prosody elicited greater neural activity compared to words produced in ADS, without prosodic exaggeration, in 13-month-old infants, supporting the mappings between word forms and their referents. Despite these findings, it appears that exaggerated prosody in IDS alone is not a significant predictor of infant language outcomes. While it facilitates performance in some experimental tasks, a meta-analysis by Spinelli and colleagues indicated no conclusive evidence that individual differences in prosodic exaggeration in naturally produced maternal IDS relate to

infants' concurrent or future linguistic skills (Spinelli et al., 2017). Therefore, it appears that the linguistic benefits demonstrated for vowel space expansion cannot be entirely attributed to the prosodic exaggeration in IDS.

This study focused on teasing apart the roles that exaggerated prosody and vowel space expansion play in facilitating early lexical development, in particular infants' ability to recognize familiar words. Only one study to date has directly contrasted the role of these two components on infants' word recognition. Song and colleagues (2010) assessed how slow speaking rate, vowel space expansion, and expanded pitch range in IDS impact lexical processing in 19-month-old infants. They compared infants' performance in typical naturally produced IDS with performance in each of three modified-IDS conditions: (i) *fast*-IDS that lacked the usual slow speaking rate, (ii) *hypo*-articulated-IDS that lacked the usual expanded vowels, and (iii) *monotonous*-IDS that lacked the usual expanded pitch range. These comparisons demonstrated that slower speaking rate significantly improved infants' lexical processing accuracy and latency. Also, the typical-IDS condition yielded shorter response latencies compared to the *hypo*-IDS condition suggesting a potentially facilitative role for vowel space expansion. However, it should be noted that the typical-IDS condition in this study consisted of a combination of the vowel space expansion and exaggerated prosody components, so it is difficult to determine whether vowel space expansion alone would be sufficient to facilitate infants' lexical processing, or whether the combination of exaggerated pitch height, pitch range, and vowel space expansion is required. Thus, the question remains as to whether vowel space expansion facilitates lexical processing *independently* of exaggerated prosody. Manipulating the presence and absence of a particular feature in both IDS and ADS may provide some answers.

This strategy was adopted in a recent study by van der Feest and colleagues (2019) who assessed word recognition in adults using three different listener-oriented speaking

styles: clear speech (ADS with vowel space expansion), IDS (IDS with vowel space expansion), and conversational speech (ADS without vowel space expansion). Adults heard these speech styles in clear listening conditions and in noise, and both offline (intelligibility) and online (response latency) measures of word recognition were collected. Results showed an overall processing advantage for both clear speech and IDS compared to conversational speech. However, nuanced differences emerged depending on the speaking style, listening condition, and word recognition measure. Offline performance in noise was superior in response to IDS than clear speech, so adults found IDS stimuli to be most intelligible. On the other hand, online performance in quiet listening conditions was superior in response to clear speech than IDS, so adults' word recognition was fastest when they heard clear speech. These findings suggest that the processing benefits of each speaking style depend on the listeners' needs and each communicative situation, but that the phonetic exaggeration properties shared between clear speech and IDS foster word recognition even in experienced adult listeners. Interestingly, however, under challenging listening conditions (in noise), IDS was most intelligible for adults, arguably because of its attention-getting prosodic properties (and possibly a novelty effect since it is not typical for adults to be addressed in IDS), which were absent in the clear and conversational speech styles. Therefore, younger infants, who are in the process of language acquisition and for whom word recognition is still a challenging task, may also require exposure to both phonetic exaggeration and prosodic cues to succeed in this task.

The present study addressed this issue by assessing 18-month-old infants' lexical access and recognition of lexical forms in two speech conditions: *Exaggerated Pitch Condition* and *Expanded Vowel Space Condition*. In the Exaggerated Pitch condition, speech consisted of acoustic exaggerations typical of IDS such as higher pitch and wider pitch range, but without vowel space expansion (*only attentional cues*). The Expanded Vowel Space

condition employed solely the phonetic exaggeration from IDS – vowel space expansion, while all other features were characteristic of ADS such as lower pitch and reduced pitch range (*only phonetic exaggeration*). We administered these conditions in a between-subjects design, similarly to Song et al. (2010), to reduce task duration and eliminate potential transfer effects from exposure to one register to another. We employed a looking-while-listening paradigm (LWL, Fernald et al., 2008) to measure infants' accuracy and latency in recognising the visual referent of a familiar word in real time. The age of 18 months was selected because it marks the time when infants' expressive vocabulary undergoes significant growth, and also it is an age when there is a significant increase in their word-recognition speed and efficiency (Fernald, 2000). Given this, the effect of infants' concurrent expressive vocabulary on lexical processing was also measured since infants with larger vocabularies were expected to show greater word recognition accuracy and shorter response latencies (Fernald et al., 2001; Zangl et al., 2005). Despite these significant gains in lexical competence, the properties of IDS to 18-month-olds, specifically vowel space expansion and pitch exaggeration have been demonstrated to remain exaggerated relative to ADS as it is found in IDS to younger infants (Kalashnikova & Burnham, 2018; Narayan & McDermott, 2016), and there is evidence that infants at this age should continue to benefit from the acoustic features of IDS in language processing tasks (Ma et al., 2011; Song et al., 2010).

Two alternative hypotheses were tested. First, if it is the case that vowel space expansion facilitates lexical processing independently from the prosodic exaggeration in IDS, then greater accuracy and shorter response latencies were expected in the Expanded Vowel Space condition (present vowel space expansion) compared to the Exaggerated Pitch condition (absent vowel space expansion) (Song et al., 2010; van der Feest et al., 2019). Alternatively, if vowel space expansion alone is not sufficient to foster word recognition performance in infants, then greater accuracy and shorter response latencies were expected in

the Exaggerated Pitch condition (present exaggerated prosody) compared to the Expanded Vowel Space condition (absent exaggerated prosody) (Graf Estes & Hurley, 2013; Thiessen et al., 2005), in which case, infants' performance would benefit from the attention-getting properties of the Exaggerated Pitch condition.

Method

Participants

Thirty-nine (16 female) full-term born monolingual Australian English-learning 18-month-old infants participated. Infants were randomly assigned to one of the between-subjects conditions: Exaggerated Pitch ($n = 19$; Age range: 17.82 - 20.71 months, $M = 18.69$, $SD = .78$), and Expanded Vowel Space ($n = 20$; Age range: 17.75 - 20.39, $M = 18.82$, $SD = .82$). Infants' age did not differ between the conditions $t(37) = .48$, $p = .63$, *Cohen's d* = .16 (95 % *CI*s: -.49, .80). An additional 6 infants were tested but excluded because they were bilingual (1), failed to capture sufficient gaze data (3), extreme fussiness (1), and equipment failure (1). All infants had normal hearing and vision. Infants were recruited via a database of families who have expressed interest to take part in infancy research at a university laboratory.

Looking While Listening Task

Stimuli and Apparatus

The auditory stimuli consisted of six words (book, car, cup, key, sheep, shoe) embedded in two carrier phrases: "Where is the *target*?" and "Look at the *target*!". These specific words were chosen based on previous LWL studies (Fernald et al., 2006; Song et al., 2010) and based on their familiarity to 18-month-old infants. In addition, the Wordbank database (Frank, et al., 2016) was used to further cross-check this word selection against lexical acquisition norms for American English (there are no available receptive norms for Australian English). This was done to ensure that all words were matched by familiarity to

avoid presenting a target and a distracter that may differ in familiarity on a given trial. Additionally, in choosing words attention was paid to ensure that the point vowels (/i, a, u/) required for the calculation of vowel space expansion were represented in the stimuli.

To create the audio stimuli, a female native speaker of Australian English (experienced in IDS production studies) was audio recorded. The speaker produced all stimuli without addressing an interlocutor and followed specific instructions for each experimental condition. To produce speech with exaggerated pitch, she was instructed to imagine that she was addressing a young infant, and to produce expanded vowel space, she was instructed to imagine that she was addressing someone who could not hear her well, so she had to speak clearly, over-enunciating the words (Lam et al., 2012).

Sixty tokens of each carrier phrase and target word combination were recorded and from these 12 tokens were selected to serve as the experimental stimuli. Stimulus selection was based on acoustic analyses of all recorded instances that consisted of extracting Mean F0 and F0 range for each phrase, and duration, F1, F2 for the vowels in each target word. The vowel F1 and F2 values were used to calculate the vowel space area for each register using the following formula (Kuhl et al., 1997): $Vowel\ area = ABS\ \frac{1}{2} \times [(F1/a/ \times (F2/i/ - F2/u/) + F1/i/ \times (F2/u/ - F2/a/) + F1/u/ \times (F2/a/ - F2/i/)]$. The vowels from the target words containing the three point vowels (car, cup, key, sheep, shoe) were used to compute the vowel space area. Since two tokens were selected per target word, this resulted in four tokens per vowel in case of vowels /a/ and /i/, and two tokens in case of /u/.

The vowel area values were 247645.4 Hz² for the Expanded Vowel Space and 154841.3 Hz² for the Exaggerated Pitch conditions (Figure 1). According to a previous speech production study by Kalashnikova, and colleagues (2017), the Expanded Vowel Space vowel area was greater than observed in natural infant-directed speech produced by Australian English female speakers, suggesting the presence of articulatory correlates of

hyperarticulated speech (i.e., articulatory effort made to exaggerate speech production).

Furthermore, it was ensured that all selected vowels differed solely based on their F1 and F2 dimensions across registers and not in length, $t(22) = -.67, p = .51$.

Next the prosodic properties of the phrases containing the words with vowels that satisfied the above criteria for the two registers were subject to analyses of pitch height and pitch range. Independent t -tests confirmed that the Exaggerated Pitch condition had significantly greater pitch height ($M = 295.42, SD = 21.47$) than Expanded Vowel Space ($M = 238.22, SD = 19.71$), $t(22) = 6.80, p < .0001$, and greater pitch range ($M = 298.35, SD = 63.62$) than Expanded Vowel Space condition ($M = 201.39, SD = 74.35$), $t(22) = 3.43, p = .002$. There were differences in overall sentence duration between the conditions ($t(21.94) = -2.75, p = .01$, Exaggerated Pitch condition: $M = 1.29, SD = .11$; Expanded Vowel Space condition: $M = 1.41, SD = .11$); however, the conditions did not differ significantly in target word duration ($t(21.90) = -.52, p = .61$, Exaggerated Pitch condition: $M = .80, SD = .16$; Expanded Vowel Space condition: $M = .83, SD = .15$) nor speech rate ($t(18.63) = 1.98, p = .06$, Exaggerated Pitch condition ($M = .77, SD = .09$), Expanded Vowel Space condition ($M = .71, SD = .06$)).

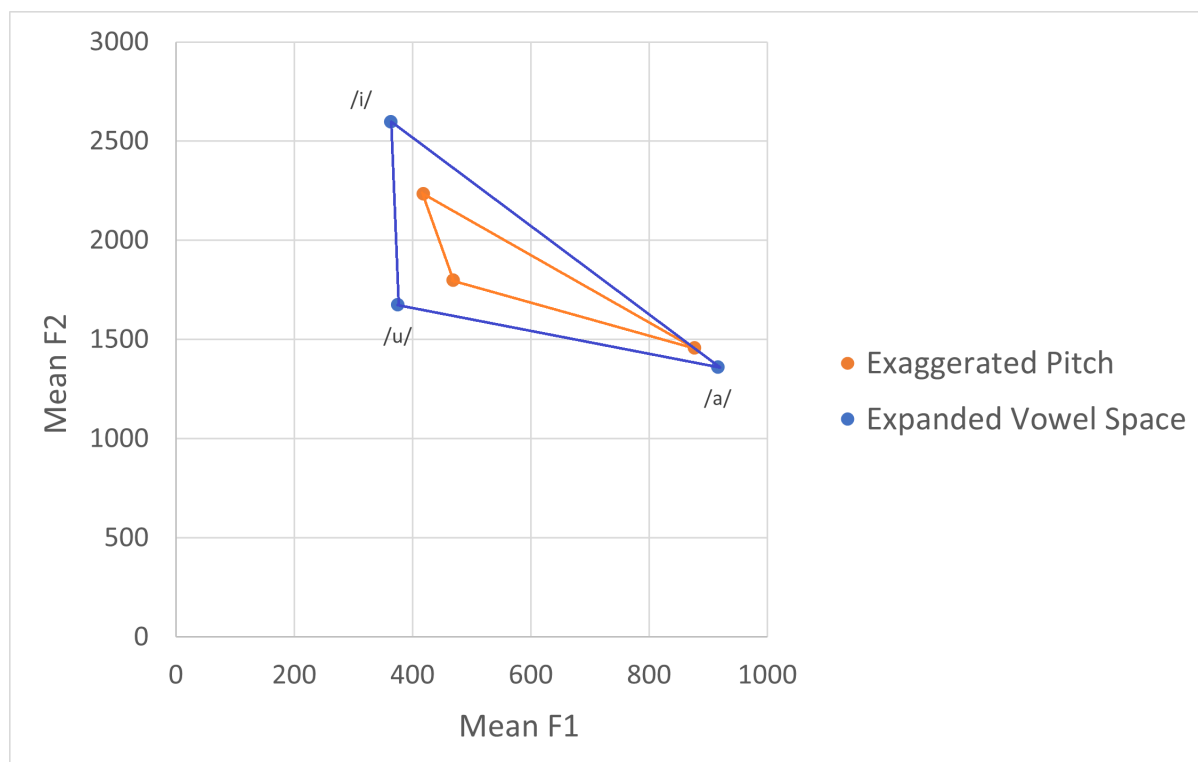


Figure 1. *Vowel space triangles constructed for stimuli used in Exaggerated Pitch and Expanded Vowel Space conditions*

The selected audio stimuli were paired with images of pairs of objects depicting the target words, 13 cm in height and separated by 18cm. Figure 2 shows an example LWL trial. The task consisted of 24 trials subdivided into 4 blocks (three trials using each of the carrier phrases) (Swingley & Aslin, 2000). The target image appeared three times on each side within each block. Two stimulus orders were created with each image presented as both the target and distracter four times. Between the blocks, filler trials were presented to maintain infants' interest. During filler trials, the images of four familiar objects were presented supplemented with sounds "look" and "wow". Images used during filler trials were not part of the test trials.

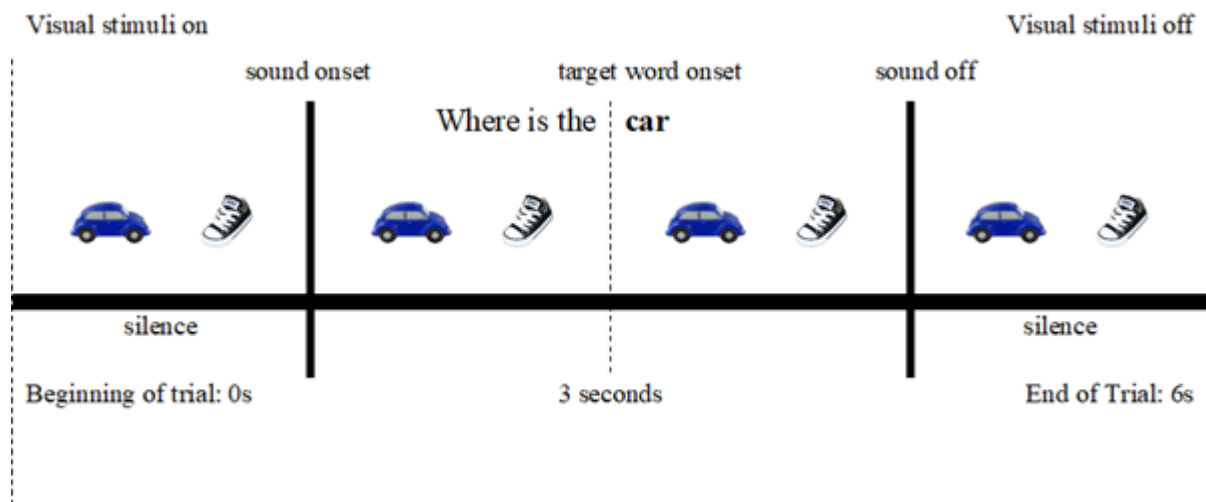


Figure 2. *Example of an experimental trial*

Stimuli were presented on a 22-inch screen using a Tobii-X120 eye tracker and Tobii Studio software to collect eye-movement data (120-Hz sampling rate). The audio stimuli were delivered through two forward-facing loudspeakers positioned below the screen.

Procedure

Infants sat on their caregiver's lap in a dimly lit soundproof laboratory room, approximately 60 cm away from the screen. Caregivers listened to masking sounds over noise-cancelling headphones and were instructed to look away from the screen to prevent their gaze from interfering with the eye-tracker's recording. At the beginning of the experiment, a 5-point infant calibration routine was completed. Before each trial, an attention-getter stimulus was presented. The experimenter observed the infant from an adjoining room and controlled the trial presentation, so each LWL trial started when the infant fixated the center of the screen.

Processing of Eye-Tracking data

Data for infants who provided less than 40% of gaze throughout the task were excluded prior to analyses (6 infants, see Participants section). The EyetrackingR package (Dink & Ferguson, 2015) in R (R Core Team, 2020) was used to process the eye-tracking data. First, two areas of interest (AOI) were defined encompassing the image of each object.

Next, two response windows were determined: the pre-naming window from 0 to 3300ms (looks prior the target label presentation) and the post-naming window from 3300 to 4800ms (looks after the target label presentation). Both response windows were a priori defined with the post-naming window being crucial for the analyses of accuracy and latency of infants' looking behavior. The 1500-ms post-naming window was determined based on previous LWL studies with similar infant ages (Brookman et al., 2020; Garrison et al., 2020; Fernald et al., 2008; Fernald & Marchman, 2012; Fernald, et al., 2016; Marchman & Fernald, 2008; Ronfard et al., 2022; Swingley & Aslin, 2002). Hence, the post-naming time window started 300ms after the target word onset to account for the time needed to process the auditory stimulus and to initiate an eye movement and finished 1800ms after target word onset since looking responses later in the trial might not represent the responses to the target but might result from irrelevant factors such as habituation (Fernald et al., 2008). Next, the amount of gaze loss in each trial was calculated and trials with over 50% gaze loss were removed. Each infant contributed on average 74% of detected gaze per trial, with no significant difference between conditions, $t(29.94) = -1.29, p = .21, \text{Cohen's } d = -.47$ (95 % CIs: -1.19, .26). Finally, the difference between the proportion of target looking out of the total looking time to the target and the distracter in pre-naming and post-naming phases (accuracy) and latency of the first look to the target object were calculated to be used as dependent variables in statistical analyses.

Vocabulary Size

Infants' caregivers completed the OZI: Australian English Communicative Development Inventory (Kalashnikova, Schwarz, & Burnham, 2016), which is the Australian English adaptation of the MacArthur-Bates Communicative Development Inventory (Fenson et al., 1994). It is a checklist consisting of 558 words that may be familiar to infants and toddlers between 12 and 30 months of age. Caregivers were required to select the words that

their child was able to understand and produce (expressive vocabulary). An independent samples *t*-test confirmed that infants' expressive vocabulary scores did not differ between the Expanded Vowel Space ($M = 92.05$, $SD = 73.71$) and the Exaggerated Pitch ($M = 112.74$, $SD = 121.25$) conditions, $t(29.42) = -.64$, $p = .53$, *Cohen's d* = $-.24$ (95 % *CI*s: $-.96$, $.49$).

Results

Figure 3 depicts the time course of infants' looking to the target object in the Exaggerated Pitch and Expanded Vowel Space conditions. We conducted two types of analysis: a window analysis (response accuracy), and an onset contingent analysis (response latency). To conduct analyse response accuracy, we first calculated the proportion of fixation time to the target object in response to hearing the target label out of the total fixation time to the target and the distracter separately in the pre-naming and the post-naming phase. These proportion values were used for comparisons against chance level (chance = $.5$). Next, in order to control for potential visual biases to the images used in this task, difference scores were computed by subtracting the proportion of looking time to the target in the pre-naming phase from the proportion of looking time to the target in the post-naming phase. These scores captured the degree to which infants' looking time to the target increased after hearing its label regarding their baseline level of attention to the target's image and were used as a measure of response accuracy to compare performance across conditions. To assess latency, that is how quickly infants switched their looking to the target object when hearing the target label, an onset contingent analysis was conducted. This analysis distinguishes between two types of trials. *Target-initial trials*, in which infants were looking at the target object at the target-label onset and *Distracter-initial trials*, in which infants were looking at the distracter object at the target-label onset. Only the *Distracter-initial trials* were of interest here since latency represents the speed of the shifts in looking away from the distracter towards the target object in response to the target label (Fernald et al., 2008). All analyses controlled for

infants' vocabulary size as a covariate given the extensive evidence that it is a significant predictor of infants' accuracy and latency in LWL tasks (Fernald et al., 2006, 2013).

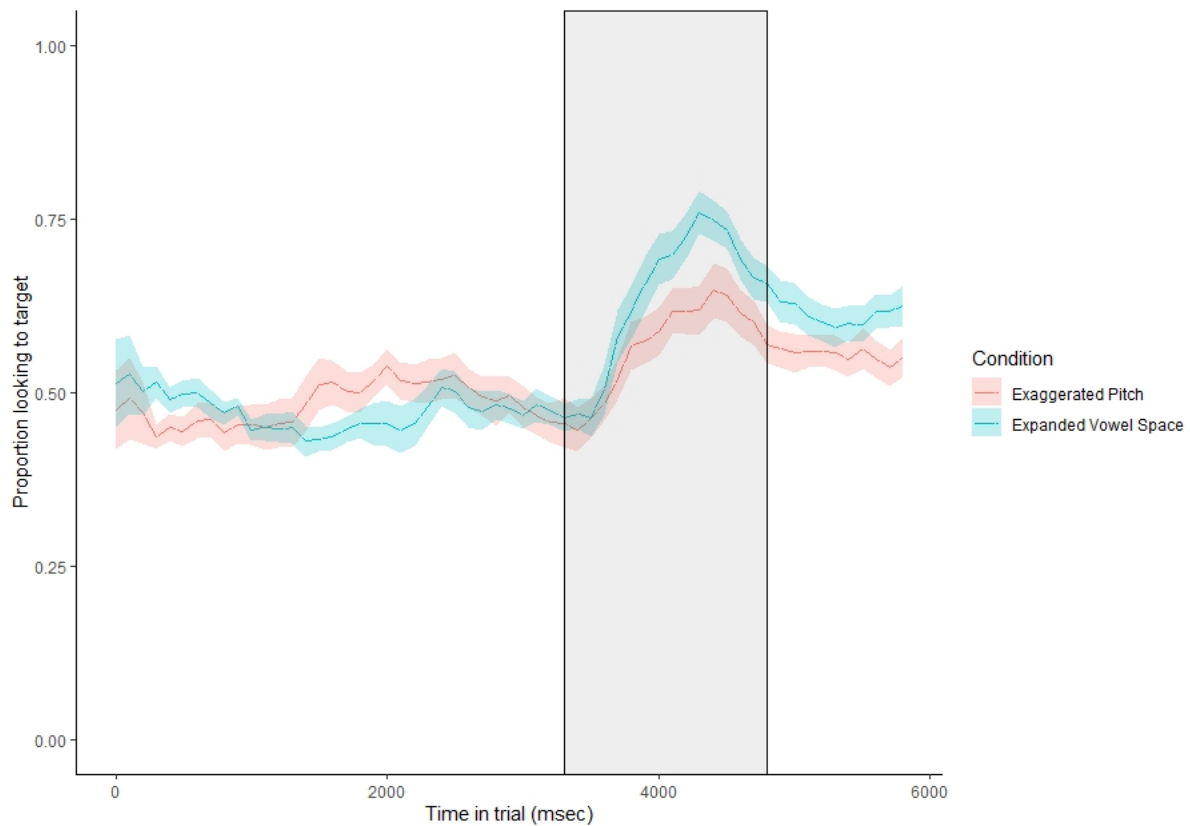


Figure 3. *Time course plot of the proportion of looking time to target across trials (samples taken every 100ms, shading around lines represents 95 % confidence intervals). Shaded rectangle area represents the post-naming analysis window (3300 - 4800ms)*

To compare accuracy between conditions, a linear mixed effects model (LME) was fitted using the *lmer* function of the *lme4* package (Bates et al., 2015) in R (R core team, 2020). The LME model was fitted with Condition and Expressive Vocabulary as the independent variables, random intercepts for Participants and Trials, and Difference score as the dependent variable. In order to compare latency between conditions, a linear model (LM) was fitted using the *lm* function of the *lme4* package (Bates, et al., 2015) in R (R core team, 2020). The LM model was fitted with Condition and Expressive Vocabulary as the independent variables and Mean Latency as the dependent variable. The significance of the

models was assessed using ANOVAs with Satterthwaite's method using the *anova* function of the *lmerTest* package (Kuznetsova et al., 2017). As a measure of effect size Cohen's *d* was calculated from the *F*-statistic using the function *F_to_d* from the package *effectsize* (Ben-Shachar et al., 2020).

Accuracy

First, one-sample *t*-test analyses (two-tailed) were conducted to compare infants' proportion of looking time to the target pre- and post-naming in the two experimental conditions to chance levels (chance = .5) (see Table 1 for *t*-tests results). In the pre-naming phase, infants in the Exaggerated Pitch condition did not differ from chance, but infants in the Expanded Vowel Space conditions looked at the target below chance level, showing an initial preference for the distracter. Performance in the Expanded Vowel Space condition was unexpected, so we conducted a further two-sample *t*-test analysis, which indicated that despite looking to the target below chance levels, infants' pre-naming target looking time did not differ significantly between the two conditions, $t(36.48) = -.96, p = .34, \text{Cohen's } d = -.32$ (95 % *CI*s: $-.97, .34$). In the post-naming phase, infants' looking times in both conditions were significantly above chance level (see Table 1 for detailed results).

The LME results using difference scores (post-naming – pre-naming) demonstrated a significant main effect of Condition, $F(1, 35.61) = 9.47, p = .004, \text{Cohen's } d = 1.03$ (95 % *CI*s: $.33, 1.72$). Infants were more accurate in the Expanded Vowel Space ($M = .17, SD = .33$) compared to the Exaggerated Pitch condition ($M = .08, SD = .30$), $p = .01$ (see Figure 4). There was no main effect of Expressive Vocabulary on performance, $F(1, 33.27) = 1.58, p = .22, \text{Cohen's } d = .44$ (95 % *CI*s: $-.25, 1.12$) (see Table 2 for the full model output).

Table 1

Results of one-sample one-tailed t-test analyses comparing the proportion of looking to the target object against chance (0.5) in the Exaggerated Pitch and Expanded Vowel Space conditions

Condition	Mean (SD)	t	df	p	Cohen's d
<i>Pre-naming phase</i>					
Exaggerated Pitch	.48 (.05)	-1.61	18	.12	-.76
Expanded Vowel Space	.47 (.04)	-3.29*	19	.01	-1.47
<i>Post-naming phase</i>					
Exaggerated Pitch	.56 (.09)	3.04**	18	.004	1.43
Expanded Vowel Space	.63 (.10)	5.65***	19	< .0001	2.59

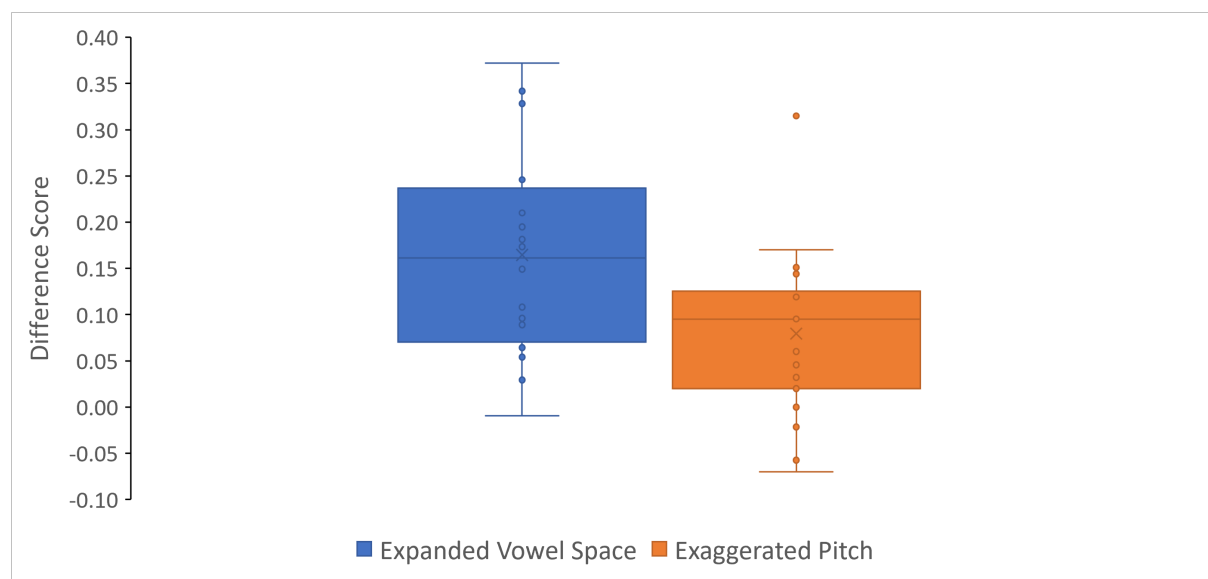


Figure 4. *Response accuracy (difference scores between the proportion of looking time to the target in the pre-naming phase and the proportion of looking time to the target in the post-naming phase) for infants in Exaggerated Pitch compared to Expanded Vowel Space condition (the circles represent individual data points, the lines represent the means, the error bars represent ± 2 SD, and the boxes encompass interquartile range)*

Table 2

The results of linear-mixed effect model for mean response accuracy between the Exaggerated Pitch and Expanded Vowel Space conditions (Condition and Expressive Vocabulary as predictors)

Fixed effects					
	<i>Estimate</i>	<i>SE</i>	<i>Df</i>	<i>t-value</i>	<i>p</i>
Intercept	.10	.03	39.29	4.01	< .0001
Condition: Clear Speech	.05	.02	35.61	3.08	.004
Expressive Vocabulary	.0002	.0002	33.27	1.26	.22
Random effects					
	<i>Variance</i>	<i>SD</i>			
Participants	.003	.06			
Trials	.004	.06			
Residuals	.09	.31			

Latency

The LM model constructed to assess response latency across conditions (Figure 5) showed no significant main effects of Condition, $F(1, 34) = .03, p = .86, \text{Cohen's } d = .06$ (95 % *CI*s: -.61, .73), and Expressive Vocabulary, $F(1, 34) = .87, p = .36, \text{Cohen's } d = .32$ (95 % *CI*s: -.36, .99).

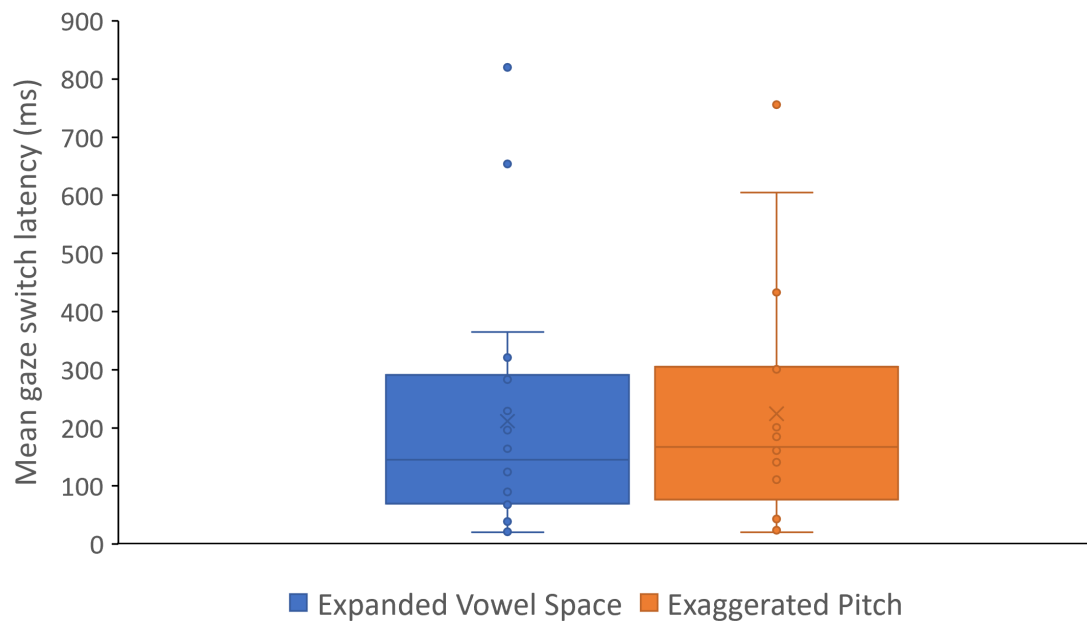


Figure 5. Mean response latency (ms) for Distracter-Initial trials in the Exaggerated Pitch and Expanded Vowel Space conditions

Discussion

This study presented 18-month-old infants with two types of speech conditions in a lexical processing task: Exaggerated Pitch (with exaggerated pitch height and pitch range, but without vowel space expansion) and Expanded Vowel Space (with vowel space expansion, but with no exaggeration in pitch height and pitch range). As would be expected for infants of this age, infants were able to recognize the highly familiar target words above chance level, and they did so with similar latency across conditions, but infants' overall accuracy was higher when they heard the words produced in the Expanded Vowel Space condition compared to the Exaggerated Pitch condition. These results suggest that vowel space expansion facilitates infants' performance on a lexical processing task even when it does not cooccur with the prosodic exaggeration that is typical of IDS.

These findings add to the growing body of evidence showing that the presence of vowel space expansion in infants' speech input yields benefits for their speech processing.

Previous studies have left open the question of whether vowel space expansion leads to these benefits independently from the prosodic cues present in IDS, as is the case of IDS heard by infants in daily interactions (Hartman, et al., 2017; Kalashnikova & Carreiras, 2021; Kalashnikova & Burnham, 2018; Liu et al., 2003; Lovcevic et al., 2020; Marklund et al., 2021), and IDS stimuli used in the previous experimental study by Song et al. (2010) that also assessed the role that IDS plays in infants' lexical processing. By using the Expanded Vowel Space and the Exaggerated Pitch conditions here, we were able to shed light on this question and demonstrate that infants' word recognition performance was higher when they heard a prosodically-neutral speech register containing words with expanded vowel spaces compared to a prosodically-exaggerated speech register that did not contain the feature of phonetic exaggeration.

These findings dovetail with extensive evidence from adult studies that vowel space expansion is one of the main correlates of clear speech. That is, more intelligible speakers produce speech with larger vowel triangle areas (Bond & Moore, 1994; Bradlow et al., 1996; Byrd, 1994; Hazan & Markham, 2004). And the presence of vowel space expansion increases speech intelligibility for listeners (Ferguson & Kewley-Port, 2007; Moon & Lindblom, 1994; Smiljanic & Bradlow, 2005), especially under challenging listening conditions (van der Feest et al., 2019). For instance, an analysis of an intelligibility database (Bradlow et al., 1996) with around 2000 sentences produced by 20 different speakers demonstrated that speech intelligibility was not related to mean pitch height or speech rate, but only to the vowel triangle area size. In other words, speakers who produced speech with larger vocalic spaces were rated as more intelligible compared to speakers producing speech with smaller vocalic spaces. These intelligibility effects are consistent across listeners' age groups: for seven-year-old children, 12-year-olds, and adults (Hazan & Markham, 2004). As we show in this study, the expanded

vowel space might have a facilitatory effect much earlier on, at 18 months of age, resulting in greater speech processing efficiency.

In their study with adult participants, van der Feest et al. (2019) showed that adults' word recognition in noise was more accurate when they heard IDS compared to a clear speech register. Recall that in that study, both registers contained expanded vowel spaces, but only IDS also contained attention-grabbing prosodic exaggeration, which was absent in clear speech. Thus, it could be that in a challenging listening condition, the attentional cues of IDS may be necessary to aid word recognition in adults. We reasoned that if this were the case, young infants acquiring language for whom word recognition is still challenging would also rely to a greater extent on prosodic exaggeration than on vowel space expansion in our stimuli and achieve greater word-recognition accuracy when hearing prosodically exaggerated speech (Exaggerated Pitch condition). However, as our accuracy results show, this was not the case. One possibility that we cannot discard is that if younger infants were to be studied on the current task, they may have shown exactly this performance pattern, hence, using the attention-getting properties of speech to aid their performance. It is important to bear in mind that the current study included 18-month-old infants, the age corresponding to the vocabulary spurt (Bloom, 1973). It is possible that for these infants the phonetic exaggeration in speech (vowel space expansion) may have more importance than attentional components of speech (pitch height and pitch range). Notably, previous findings suggest that infants during their second year are able to process successfully speech registers other than IDS. For example, Ma and colleagues (2011) showed that 21-month-old infants with larger expressive vocabularies were successful in learning words from ADS suggesting that greater lexical competence reduces infants' dependence on the prosodic exaggeration in IDS. However, it is noteworthy that this effect found for word learning may not generalize to the present word recognition task where infants' recognition of highly familiar words was

assessed. It remains of interest for future studies to systematically assess infants' reliance on the phonetic and acoustic features of IDS across ages and across a variety of language processing tasks.

An unexpected finding in this study is that the benefit of Expanded Vowel Space over Exaggerated Pitch was observed solely in infants' response accuracy and not response latency. Several possibilities could account for this result. First, this pattern is the same as that in the van der Feest et al. (2019) study, in which it was found that adults also showed more accurate but not faster responses when presented with IDS compared to clear speech. The authors suggested that adults were equally quick to orient their gaze to the target because of their high lexical processing efficiency, but this is unlikely to be the case for 18-month-old infants. Instead, it is likely that response latency was not a sensitive measure for capturing the effects of speech register on infants' performance in our task. Such a possibility is supported by the results of similar studies employing the LWL method to assess infants' lexical processing. Specifically, Fernald and colleagues (Fernald et al., 2006) demonstrated that between 15 and 18 months of age, there is wide and unstable discrepancy in the latency of lexical processing. Accordingly, several studies have reported discrepant results for accuracy and latency measures in LWL tasks (e.g., Brookman et al., 2020; Suttora et al., 2017). We also acknowledge the possibility that this discrepancy may be due to our unexpected finding of a preference for the distracter in the pre-naming phase among infants in the Vowel Space Expansion condition. This preference may have led to greater delays in disengaging from the distracter, thus obscuring possible advantages in lexical processing related to the phonetic exaggeration of stimuli used in this condition. Unfortunately, our design does not allow us to directly test this possibility; however, it remains remarkable that despite this initial preference for the distracter, infants increased their fixations to the target, significantly more than

chance, in the Vowel Space Expansion condition compared to the Exaggerated Pitch condition.

It is of course noteworthy that the acoustic manipulations in this study do not reflect the acoustic properties of infants' daily linguistic input. We created an Expanded Vowel Space condition, which corresponded to hyperarticulated adult-directed speech. However, there is evidence that the exaggerated acoustic properties of IDS emerge from different articulatory adjustments that consist of laryngeal raising and not actual *hyper-articulation* (Kalashnikova et al., 2017). Laryngeal raising raises all formant frequencies, including fundamental frequency. Thus, in their everyday experience, infants do not hear speech with only clarity or only attentional cues, but rather they hear speech with both cues combined. This combination presents a challenge for research trying to disentangle the role of each IDS component in fostering infants' attentional, emotional, or language processing skills, as it was the case in this study. Despite this limitation, we argue that our results show that exposure to exaggerated prosodic properties alone (Exaggerated Pitch condition) is not sufficient to boost performance to the same degree as exposure to exaggerated phonetic properties alone (Expanded Vowel Space condition). While this finding provides evidence for benefits of vowel space expansion in IDS in infants' language processing, it remains most likely that a combination of prosodic and phonetic exaggeration in this register is most conducive to successful language processing and acquisition.

It should be noted that these findings might not generalize across different infant populations due to potential differences in both quantity and quality of IDS. Indeed, previous evidence suggests differences in quantity of IDS across different cultures with infants in non-Western cultures hearing less IDS compared to infants from Western cultures (Cristia et al., 2019; Shneidman & Goldin-Meadow, 2012; Weber et al., 2017). Additionally, the quantity of IDS has been found to vary across different cultures within the same society (Farran et al.,

2016) as well as across populations from different SES backgrounds (Hart & Risley, 1995). Furthermore, the quality of IDS has been found to vary as well. For example, vowel space expansion has not been found in Danish, German, Dutch, and Norwegian (Audibert & Falk, 2018; Benders, 2013; Cox et al., 2022; Englund, 2018). Regarding the universality of exaggerated pitch features in IDS, the evidence suggests that exaggerated pitch features might be consistent across different languages and cultures. For example, exaggerated pitch features have been found in American English (Fernald & Simon, 1984), Australian English (Kalashnikova & Burnham, 2018), French, Italian, German (Fernald, 1989), Mandarin (Grieser & Kuhl, 1988), Cantonese (Rattanasone et al., 2013), Thai (Kitamura et al., 2002), and Japanese (Fernald, 1989). Even when directly comparing different cultures, Broesch et al. (2015) demonstrated exaggerated pitch height in IDS in Kenya, North America, and Fiji with no difference between cultures. So far only one study did not find evidence for exaggerated pitch in IDS. Specifically, higher pitch was absent in IDS of Quiche-speaking mothers (Bernstein Ratner & Pye, 1984), possibly because high pitch is reserved for addressing social superiors in their culture. Nevertheless, a recent corpus study (Moser et al., 2020) that assessed acoustic features of IDS and ADS across 21 urban, rural, and small-scale societies demonstrated the robust presence of the exaggerated pitch features in IDS regardless of culture. Hence, it is possible that specific IDS features might have different weights for infants across different cultures potentially resulting in different relations of specific features to certain linguistic outcomes.

Another factor that might affect infants' processing of specific IDS features is infants' age. It is possible that at different ages infants might process certain IDS features differently depending on their stage of language acquisition. Finally, infants' processing of IDS features might depend on infants' preference for certain IDS features at specific ages. Indeed, previous evidence suggests that infants' preference for specific IDS features changes across

age. For example, it has been found that four-to six-month-old infants preferer IDS with stretched vowels, slow tempo, and high positive affect, while 10-month-olds prefer IDS with normal vowel duration and normal tempo regardless of affect (Kitamura & Notley, 2009; Panneton et al., 2006). These differences in infants' preference for specific IDS features may result from their emerging developmental needs. Thus, it is possible that infants' preference for IDS features investigated in this study might change resulting in different attention to these features during speech processing. Hence, future studies should investigate infants' use of specific IDS features across different ages and across different cultures. This will aid our understanding of specific IDS features and their effects on infants' language development.

Over the last few decades, evidence from infancy research has suggested that adults' speech to infants, IDS, with its exaggerated features positively affects infants' emotional and cognitive development and promotes infants' language acquisition. However, the role of particular individual features of IDS in this process has been less indisputable. The purpose of the current study was to determine the independent effects of phonetic exaggeration, and prosodic exaggeration, in IDS on 18-month-old infants' speech processing. Our findings underline the facilitative role of phonetic exaggeration in retrieval and recognition of lexical forms in 18-month-old infants, even when being exposed to speech that lacks the attention-getting properties of typical IDS. Taken together, our findings indicate that even though phonetic exaggeration might not be a universal feature of IDS, it may play a linguistic function independently of other IDS features.

References

- Adriaans, F., & 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. <https://doi.org/10.1121/1.4982246>
- Audibert, N., & Falk, S. (2018, June). Vowel space and f0 characteristics of infant-directed singing and speech. In *Proceedings of the 19th international conference on speech prosody* (pp. 153-157).
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Benders, T. (2013). Mommy is only happy! Dutch mothers' realization of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development*, *36*(4), 847-862. <https://doi.org/10.1016/j.infbeh.2013.09.001>
- Benders, T., StGeorge, J., & Fletcher, R. (2021). Infant-directed speech by Dutch fathers: Increased pitch variability within and across utterances. *Language Learning and Development*, *17*(3), 292-325. <https://doi.org/10.1080/15475441.2021.1876698>
- Bloom, L. (1973). *One word at a time: the use of single-word utterances before syntax*. The Hague: Mouton
- Boersma, P. P. G., & Weenink, D. J. M. (1996). *Praat: Doing Phonetics by Computer: Version 3.4*. Instituut voor Fonetische Wetenschappen.
- Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech communication*, *14*(4), 325-337. [https://doi.org/10.1016/0167-6393\(94\)90026-4](https://doi.org/10.1016/0167-6393(94)90026-4)

- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech communication, 20*(3-4), 255-272. [https://doi.org/10.1016/S0167-6393\(96\)00063-5](https://doi.org/10.1016/S0167-6393(96)00063-5)
- Brookman, R., Kalashnikova, M., Conti, J., Xu Rattanasone, N., Grant, K. A., Demuth, K., & Burnham, D. (2020). Maternal depression affects infants' lexical processing abilities in the second year of life. *Brain sciences, 10*(12), 977. <https://doi.org/10.3390/brainsci10120977>
- Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435-1435.
- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software, 80*(1), 1-28. <http://dx.doi.org/10.18637/jss.v080.i01>
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication, 15*(1-2), 39-54. [https://doi.org/10.1016/0167-6393\(94\)90039-6](https://doi.org/10.1016/0167-6393(94)90039-6)
- Castellanos, A., Benedí, J. M., & Casacuberta, F. (1996). An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Communication, 20*(1-2), 23-35. [https://doi.org/10.1016/S0167-6393\(96\)00042-8](https://doi.org/10.1016/S0167-6393(96)00042-8)
- Chong, S. C. F., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development: An International Journal of Research and Practice, 12*(3), 211-232. <https://doi.org/10.1002/icd.286>
- Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass, 7*(3), 157-170. <https://doi.org/10.1111/lnc3.12015>

- Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913-934.
<https://doi.org/10.1017/S0305000912000669>
- Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child development*, *90*(3), 759-773. <https://doi.org/10.1111/cdev.12974>
- Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child development*, *61*(5), 1584-1595. <https://doi.org/10.1111/j.1467-8624.1990.tb02885.x>
- Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2022). A Systematic Review and Bayesian Meta-Analysis of the Acoustic Features of Infant-Directed Speech. <https://doi.org/10.21203/rs.3.rs-1396005/v1>
- Cox, C. M. M., Dideriksen, C., Keren-Portnoy, T., Roepstorff, A., Christiansen, M. H., & Fusaroli, R. (2022). Infant-Directed Speech Does Not Always Involve Exaggerated Vowel Distinctions: Evidence From Danish. <https://doi.org/10.31234/osf.io/2gswt>
- Dink, J. W., & Ferguson, B. (2015). eyetrackingR: An R library for eye-tracking data analysis. Available at www.eyetracking-r.com. Accessed April, 6, 2019.
- Dunst, C., Gorman, E., & Hamby, D. (2012). Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning*, *5*(1), 1-13.
- Englund, K., & Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development: An International Journal of Research and Practice*, *15*(2), 139-160. <https://doi.org/10.1002/icd.445>
- Englund, K. T. (2018). Hypoarticulation in infant-directed speech. *Applied Psycholinguistics*, *39*(1), 67-87. <https://doi.org/10.1017/S0142716417000480>

- Erickson, D. (2002). Articulation of extreme formant patterns for emphasized vowels. *Phonetica*, 59(2-3), 134-149. <https://doi.org/10.1159/000066067>
- Farran, L. K., Lee, C. C., Yoo, H., & Oller, D. K. (2016). Cross-cultural register differences in infant-directed speech: An initial study. *PloS one*, 11(3), e0151518. <https://doi.org/10.1371/journal.pone.0151518>
- Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental psychology*, 20(1), 104. <http://dx.doi.org/10.1037/0012-1649.20.1.104>
- Fernald, A. (2000). Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition. *Phonetica*, 57(2-4), 242-254. <https://doi.org/10.1159/000028477>
- Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child development*, 64(3), 657-674. <https://doi.org/10.1111/j.1467-8624.1993.tb02934.x>
- Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental psycholinguistics: On-line methods in CHILDREN'S LANGUAGE PROCESSING*, 44, 97.
- Fernald, A., Swingle, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child development*, 72(4), 1003-1015. <https://doi.org/10.1111/1467-8624.00331>
- Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, 42(1), 98. <https://psycnet.apa.org/doi/10.1037/0012-1649.42.1.98>

- Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental science, 16*(2), 234-248. <https://doi.org/10.1111/desc.12019>
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of child language, 1*-18. doi: [10.1017/S0305000916000209](https://doi.org/10.1017/S0305000916000209)
- Garrison, H., Baudet, G., Breitfeld, E., Aberman, A., & Bergelson, E. (2020). Familiarity plays a small role in noun comprehension at 12–18 months. *Infancy, 25*(4), 458-477. <https://doi.org/10.1111/infa.12333>
- Gergely, A., Faragó, T., Galambos, Á., & Topál, J. (2017). Differential effects of speech situations on mothers' and fathers' infant-directed and dog-directed speech: An acoustic analysis. *Scientific reports, 7*(1), 1-10.
- Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy, 18*(5), 797-824. <https://doi.org/10.1111/infa.12006>
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.
- Hartman, K. M., Ratner, N. B., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of child language, 44*(5), 1140-1162. <https://doi.org/10.1017/S0305000916000520>
- Hazan, V., Uther, M., & Grunlund, S. (2015). How does foreigner-directed speech differ from other forms of listener-directed clear speaking styles?.
- Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America, 116*(5), 3108-3118. <https://doi.org/10.1121/1.1806826>

- Han, M., De Jong, N., & Kager, R. (2020). Pitch properties of infant-directed speech specific to word-learning contexts: A cross-linguistic investigation of Mandarin Chinese and Dutch. *Journal of Child Language*, 47(1), 85-111.
<https://doi.org/10.1017/S0305000919000813>
- Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., ... & Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, 1-12.
- Kalashnikova, M., Carignan, C., & Burnham, D. (2017). The origins of babytalk: smiling, teaching or social convergence?. *Royal Society open science*, 4(8), 170306.
<https://doi.org/10.1098/rsos.170306>
- Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of child language*, 1-19. <https://doi.org/10.1017/S0305000917000629>
- Kalashnikova, M., Schwarz, I. C., & Burnham, D. (2016). OZI: Australian English communicative development inventory. *First Language*, 36(4), 407-427.
<https://doi.org/10.1177%2F0142723716648846>
- Kalashnikova, M., & Carreiras, M. (2021). Input quality and speech perception development in bilingual infants' first year of life. *Child Development*.
<https://doi.org/10.1111/cdev.13686>
- Kaplan, P. S., Jung, P. C., Ryther, J. S., & Zarlengo-Strouse, P. (1996). Infant-directed versus adult-directed speech as signals for faces. *Developmental Psychology*, 32(5), 880.
<http://dx.doi.org/10.1037/0012-1649.32.5.880>
- Kitamura, C., & Notley, A. (2009). The shift in infant preferences for vowel duration and pitch contour between 6 and 10 months of age. *Developmental science*, 12(5), 706-714.
<https://doi.org/10.1111/j.1467-7687.2009.00818.x>

- 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. (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>
- Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*.
[https://doi.org/10.1044/1092-4388\(2012/11-0154\)](https://doi.org/10.1044/1092-4388(2012/11-0154))
- Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental science*, 6(3), F1-F10.
<https://doi.org/10.1111/1467-7687.00275>
- Lorge, I., & Katsos, N. (2019). Listener-adapted speech: Bilinguals adapt in a more sensitive way. *Linguistic Approaches to Bilingualism*, 9(3), 376-397.
<https://doi.org/10.1075/lab.16054.lor>
- Lovcevic, I., Kalashnikova, M., & Burnham, D. (2020). Acoustic features of infant-directed speech to infants with hearing loss. *The Journal of the Acoustical Society of America*, 148(6), 3399-3416. <https://doi.org/10.1121/10.0002641>
- Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development*, 7(3), 185-201.
<https://doi.org/10.1080/15475441.2011.579839>
- 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%2F2515245919900809>

- Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science*, *11*(3), F9-F16. <https://doi.org/10.1111/j.1467-7687.2008.00671.x>
- Marklund, E., Marklund, U., & Gustavsson, L. (2021). An Association Between Phonetic Complexity of Infant Vocalizations and Parent Vowel Hyperarticulation. *Frontiers in Psychology*, *2873*. <https://doi.org/10.3389/fpsyg.2021.693866>
- Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological science*, *26*(3), 341-347. <https://doi.org/10.1177/0956797614562453>
- McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence?. *Cognition*, *129*(2), 362-378. <https://doi.org/10.1016/j.cognition.2013.07.015>
- Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, *251*, 118991. <https://doi.org/10.1016/j.neuroimage.2022.118991>
- Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition*, *166*, 84-93. <https://doi.org/10.1016/j.cognition.2017.05.003>
- Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical society of America*, *96*(1), 40-55. <https://doi.org/10.1121/1.410492>

- Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language*, 22(2), 171-184.
<https://doi.org/10.1016/j.csl.2007.07.003>
- Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J. I., & Shozo, K. (2012). Cerebral responses to infant-directed speech and the effect of talker familiarity. *Neuroimage*, 59(2), 1735-1744.
<https://doi.org/10.1016/j.neuroimage.2011.07.093>
- Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America*, 139(3), 1272-1281.
- Panneton, R., Kitamura, C., Mattock, K., & Burnham, D. (2006). Slow speech enhances younger but not older infants' perception of vocal emotion. *Research in Human Development*, 3(1), 7-19. https://doi.org/10.1207/s15427617rhd0301_2
- Peter, V., Kalashnikova, M., Santos, A., & Burnham, D. (2016). Mature neural responses to infant-directed speech but not adult-directed speech in pre-verbal infants. *Scientific reports*, 6, 34273.
- Piazza, G., Martin, C. D., & Kalashnikova, M. (2021). The acoustic features and didactic function of Foreigner Directed Speech: A literature review.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.r-project.org/>
- Rattanasone, N. X., Burnham, D., & Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *Journal of Phonetics*, 41(5), 332-343. <https://doi.org/10.1016/j.wocn.2013.06.001>

- Remick, H. (1976). *Maternal speech to children during language acquisition*. – W. von RafflerEngel, Y. Lebrun (Eds.). *Baby Talk and Infant Speech*. Amsterdam: Swets; Zeitlinger B. V., 223–233.
- Ronfard, S., Ran, W. E. I., & Rowe, L. M. (2022). Exploring the linguistic, cognitive, and social skills underlying lexical processing efficiency as measured by the looking-while-listening paradigm. *Journal of Child Language*, *49*(2), 302-325.
<https://doi.org/10.1017/S0305000921000106>
- Saito, Y., Aoyama, S., Kondo, T., Fukumoto, R., Konishi, N., Nakamura, K., ... & Toshima, T. (2007). Frontal cerebral blood flow change associated with infant-directed speech. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *92*(2), F113-F116. <http://dx.doi.org/10.1136/adc.2006.097949>
- Santesso, D. L., Schmidt, L. A., & Trainor, L. J. (2007). Frontal brain electrical activity (EEG) and heart rate in response to affective infant-directed (ID) speech in 9-month-old infants. *Brain and Cognition*, *65*(1), 14-21. <https://doi.org/10.1016/j.bandc.2007.02.008>
- Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech?. *Developmental science*, *15*(5), 659-673.
<https://doi.org/10.1111/j.1467-7687.2012.01168.x>
- Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, *14*(6), 654-666.
- Soderstrom, M., Blossom, M., Foygel, I., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, *35*, 869–902. <https://doi.org/10.1017/S0305000908008763>
- Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, *128*(1), 389-400. <https://doi.org/10.1121/1.3419786>

- Spinelli, M., & Mesman, J. (2018). The regulation of infant negative emotions: The role of maternal sensitivity and infant-directed speech prosody. *Infancy*, 23(4), 502-518.
<https://doi.org/10.1111/infa.12237>
- Spinelli, M., Fasolo, M., & 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. <https://doi.org/10.1016/j.dr.2016.12.001>
- Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of child language*, 10(1), 1-15.
<https://doi.org/10.1017/S0305000900005092>
- Stern, D. N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, 18(5), 727.
<https://psycnet.apa.org/doi/10.1037/0012-1649.18.5.727>
- Suttora, C., Salerni, N., Zanchi, P., Zampini, L., Spinelli, M., & Fasolo, M. (2017). Relationships between structural and acoustic properties of maternal talk and children's early word recognition. *First Language*, 37(6), 612-629.
<https://doi.org/10.1177/0142723717714946>
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147-166. [https://doi.org/10.1016/S0010-0277\(00\)00081-0](https://doi.org/10.1016/S0010-0277(00)00081-0)
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological science*, 13(5), 480-484.
<https://doi.org/10.1111/2F1467-9280.00485>
- Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal*

of the Acoustical Society of America, 142(2), 493-503.

<https://doi.org/10.1121/1.4995998>

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

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion?. *Psychological science*, 11(3), 188-195.

<https://doi.org/10.1111/1467-9280.00240>

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, 9(2), 335-340.

van der Feest, S. V., Blanco, C. P., & Smiljanic, R. (2019). Influence of speaking style adaptations and semantic context on the time course of word recognition in quiet and in noise. *Journal of Phonetics*, 73, 158-177. <https://doi.org/10.1016/j.wocn.2019.01.003>

Wang, L., Kalashnikova, M., Kager, R., Lai, R., & Wong, P. (2021). Lexical and Prosodic Pitch Modifications in Cantonese Infant-directed Speech. *Journal of Child Language*, 48(6), 1235-1261. <https://doi.org/10.1017/S0305000920000707>

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological science*, 25(7), 1314-1324. <https://doi.org/10.1177%2F0956797614531023>

Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: Effectiveness of a parenting program in rural Senegal. *Child Development*, 88(5), 1513-1526. <https://doi.org/10.1111/cdev.12882>

Weirich, M., & Simpson, A. (2019). Effects of gender, parental role, and time on infant-and adult-directed read and spontaneous speech. *Journal of Speech, Language, and Hearing Research*, 62(11), 4001-4014. https://doi.org/10.1044/2019_JSLHR-S-19-0047

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

https://doi.org/10.1207/s15327078in1101_2

Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development*, *6*(2), 179-208.

https://doi.org/10.1207/s15327647jcd0602_2

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., ... & Carney, E.

(2011). Neural coding of formant-exaggerated speech in the infant

brain. *Developmental science*, *14*(3), 566-581. [https://doi.org/10.1111/j.1467-](https://doi.org/10.1111/j.1467-7687.2010.01004.x)

[7687.2010.01004.x](https://doi.org/10.1111/j.1467-7687.2010.01004.x)