

Article

Not peer-reviewed version

Developmental Changes and Sex Differences in Infants' Neural Sensitivity to Emotional Prosodies in Spoken Words

[Chieh Kao](#) and [Yang Zhang](#) *

Posted Date: 3 October 2023

doi: 10.20944/preprints202309.2021.v1

Keywords: emotional prosody; mismatch response; sex differences; infancy



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Developmental Changes and Sex Differences in Infants' Neural Sensitivity to Emotional Prosodies in Spoken Words

Chieh Kao^{1,2} and Yang Zhang^{1,3}

¹ Department of Speech-Language-Hearing Sciences, University of Minnesota, Twin Cities, MN, USA; Chieh Kao <https://orcid.org/0000-0003-0154-650X>; Yang Zhang <https://orcid.org/0000-0001-6777-3487>

² Center for Cognitive Sciences, University of Minnesota, Twin Cities, MN, USA

³ Masonic Institute for the Developing Brain, University of Minnesota, Twin Cities, MN, USA

* Correspondence: Yang Zhang, University of Minnesota, 164 Pillsbury Dr SE, Minneapolis, MN 55455, USA. Email: zhanglab@umn.edu

Abstract: This study examined infants' neural responses to emotional prosody in natural speech. A multi-feature oddball paradigm was used with 34 3~11 month-old infants. Results showed distinct early (100-200 ms) and late (300-500 ms) mismatch responses to different emotional prosodies. Older infants had more negative early responses with happy and angry prosodies evoking stronger responses compared to sad prosody while younger infants showed a clearer distinction between angry and sad prosodies. In the late time window, angry prosody elicited more negative responses than sad prosody with younger infants showing more distinct responses between these two prosodies. Males exhibited stronger early mismatch responses than females. These findings call for further research on the implications for socio-emotional and language development.

Keywords: emotional prosody; mismatch response (MMR); sex differences; infancy

Introduction

Emotional prosody plays a vital role in infants' early development. It serves as a crucial cue for social interaction, bonding with caregivers, and behavior regulation (Grossmann, 2010; Mumme et al., 1996; Vaish & Striano, 2004). Within the socio-emotional framework (Hohenberger, 2011), emotional prosody directs infants' attention to relevant speech input, which may facilitate word segmentation, joint attention and turn-taking, and the interplay between social and emotional factors is critical in shaping the overall language acquisition journey. Before the age of one, infants may rely more on the paralinguistic prosodic cues than the linguistic ones (Fernald, 1989, 1993; Lawrence & Fernald, 1993). However, the mechanisms and developmental trajectories underlying infants' decoding of vocal emotions remain unclear.

Behavioral assessment of infants' voice discrimination has typically been carried out using the habituation or preferential looking paradigm (Aslin, 2007; Colombo & Mitchell, 2009; Groves & Thompson, 1970; Kao & Zhang, 2022). However, these behavioral data can only tell if infants behaviorally show distinct responses to the sounds. A failure to show increased attention to the new vocal emotional category does not necessarily mean that infants cannot tell the two vocal expressions. Instead, perceptually irrelevant factors (e.g., inherent preference for particular stimuli) may affect infants' behavioral reactions to different voices (Oakes, 2010). Therefore, a more sensitive measure is needed to capture infants' early emotional voice discrimination ability.

Electroencephalography (EEG) recordings have been proven to be an ideal tool to complement behavioral assessment (De Haan, 2007; Hartkopf et al., 2019). EEG captures neural activities at the millisecond scale and allows for the examination of event-related potentials (ERPs), which are averaged neural responses time-locked to specific stimuli, such as speech sounds. By including infants' ERP responses to emotional prosodies, researchers can further address young infants' inconsistent behavioral discrimination of emotional speech and better understand the age-dependent

changes in decoding the crucial affective information in speech that influences their socio-emotional and cognitive development (Hohenberger, 2011).

Previous developmental research on emotion cognition has predominantly focused on facial expressions rather than vocal expressions (Grossmann, 2010; Morningstar et al., 2018). Consequently, little is known about how infants perceive different emotional prosody categories in spoken words, which likely exhibits age-related and category-specific changes, similar to their visual responsiveness to facial expressions (Leppänen et al., 2007; Nelson & De Haan, 1996). Among the auditory ERP components, the mismatch response (MMR) is an important neural marker for infants' pre-attentive discrimination (de Haan, 2002; Kushnerenko et al., 2013). MMR is typically elicited by presenting infrequent sounds (deviants) within a continuous stream of frequent sounds (standards) (Näätänen et al., 2007). The time window and polarity of MMRs in infants vary depending on stimuli and age (Cheour et al., 2002; Csibra et al., 2008; Friederici et al., 2002; Maurer et al., 2003). Unlike adults, infants' MMR is often observed as a slow positive wave in a later window around 200 to 450 ms post-stimulus, and numerous studies have successfully recorded infants' MMRs to different speech and voice stimuli (e.g., Cheng et al., 2012; Dehaene-Lambertz, 2000; Friederici et al., 2002; García-Sierra et al., 2021; He et al., 2009; Leppänen et al., 2004; Shafer et al., 2012; Winkler et al., 2003).

Although limited research exists, there is evidence that infants under the age of one already exhibit different neural activities in response to different vocal emotions (Blasi et al., 2011; Grossmann et al., 2005; Kok et al., 2014; Minagawa-Kawai et al., 2011; Morningstar et al., 2018). MMR studies have even documented differential neural sensitivities to vocal emotions in newborns (Cheng et al., 2012; Kostilainen et al., 2018; Kostilainen et al., 2020; Zhang et al., 2014). For example, newborns' MMRs were elicited by sad and angry voices around 100 to 200 ms after sound onset (Kostilainen et al., 2018), and their MMRs to angry, happy, and fearful sounds differed around 300 to 500 ms post-stimulus, with a stronger MMR for happy compared to angry sounds (Kostilainen et al., 2020; Cheng et al., 2012; Zhang et al., 2014).

One issue with the previous MMR studies is that the stimuli tended to use simple syllables to carry the vocal emotional expression. Another is the mixed results regarding sex differences, which may have age-dependent and emotion-specific effects. Kostilainen et al. (2018) found that female newborns exhibited stronger MMRs to happy voices (100-200 ms) compared to males, while the opposite pattern was observed for sad voices (males showing stronger responses). However, other newborn studies did not find a sex effect on emotional MMR in a later time window (Cheng et al., 2012; Kostilainen et al., 2018). Given that sex effects have been found in adult MMR studies (Hung & Cheng, 2014; Schirmer et al., 2005), it is crucial to investigate the developmental trajectory of the sex effect on infants' pre-attentive neural processing of emotional prosody in natural speech.

The current study aims to investigate infants' ability to automatically extract different emotional prosodies in non-repeating spoken words and how this ability develops in relation to age and sex. The experimental protocol adopted a roving multi-feature oddball paradigm to examine infants' neural sensitivities to multiple types of emotional sounds (happy, angry, and sad) in a single session (Kao & Zhang, 2023). A key modification in this protocol was the use of non-repeating spoken words to deliver emotional prosody, aiming to investigate if the listener's pre-attentive neural system can detect emotional categories based on perceptual grouping of statistical regularities in acoustic cues for each emotion category. We expected to observe mismatch responses (MMRs) in early (100-200 ms) and late windows (300-500 ms) after sound onset. Category-specific differences were also hypothesized, particularly stronger late MMRs to anger, aligning with the automatic orienting response to threat-related signals observed in previous studies. Additionally, potential sex differences in infants' auditory emotional MMRs were anticipated. It was hypothesized that analogous to their superior facial expression processing abilities, female infants might exhibit more salient MMRs compared to males.

Method

Participants

A total of 46 typically developing infants from 3 to 12 months of age (male = 25, female = 21; mean age = 7 months and 15 days) were recruited through advertisements, words of mouth, and the local institution's subject pool. All infants were born full-term (38 – 42 weeks), healthy with normal hearing, and from English-speaking families. Twelve infants' data were not included in the analysis due to EEG cap being pulled off ($n = 1$), equipment failure ($n = 1$). The final sample included 34 infants between the ages of two months 26 days and 11 months 11 days (male = 18, female = 16; mean age = 7 months and 26 days). The study was approved by the Institutional Review Board. Parents signed the informed consent for their children prior to the participation and received \$20 as monetary compensation upon completion.

Stimuli

All speech stimuli were from the Toronto Emotional Speech Set (Dupuis & Pichora-Fuller, 2010), including 200 monosyllabic phonetically balanced words (Northwestern University Auditory Test No. 6, NU-6; see Table II in Tillman & Carhart, 1966). Each of the 200 words was spoken in neutral, happy, sad, and angry voices by a young female speaker, yielding a total of 800 stimuli. The sounds were sampled at 24, 414 Hz, with the mean sound intensity levels equalized using PRAAT 6.0.40 (Boersma & Weenink, 2020). Table 1 summarizes five key acoustic measures of the speech stimuli, which are commonly analyzed to characterize different vocal emotions (Amorim et al., 2019; Banse & Scherer, 1996; Johnstone & Scherer, 2000; Mani & Pätzold, 2016).

Table 1. The acoustic properties of each emotional prosody.

| Emotions | Mean F0 (Hz) | | Duration (ms) | | Intensity Variation (dB) | | HNR (dB) | | Spectral Centroid (Hz) | |
|----------|--------------|-----------|---------------|-----------|--------------------------|-----------|----------|-----------|------------------------|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Angry | 216.71 | 36.64 | 646 | 109 | 11.15 | 3.74 | 9.22 | 5.01 | 1810.96 | 1075.32 |
| Happy | 226.13 | 10.86 | 742 | 91 | 10.82 | 4.12 | 17.53 | 3.77 | 1052.92 | 265.19 |
| Sad | 180.42 | 20.57 | 822 | 104 | 10.18 | 3.11 | 19.31 | 4.25 | 408.79 | 278.89 |
| Neutral | 195.04 | 9.25 | 667 | 84 | 9.14 | 3.71 | 18.75 | 4.43 | 758.43 | 220.34 |

Note. The averaged values and standard deviations of all the words were used to report the mean fundamental frequency (F0), word duration, intensity variation, harmonics-to-noise ratio (HNR), and spectral centroid of each emotional prosody.

Procedure

The study adopted a multi-feature oddball paradigm (or optimum paradigm; Näätänen et al., 2004; Thönnessen et al., 2010) to examine infants' early neural sensitivities to happy, angry, and sad prosodies as against the neutral prosody. The multi-feature oddball paradigm is a passive listening protocol that allows us to measure three emotional prosody contrasts (from neutral to angry, from neutral to happy, and from neutral to sad) within the same recording session, which is suitable for infant participants. There were 600 trials in total. The *Standard* stimuli were words in a neutral tone (presented with 50% probability, 300 trials). The three emotional tones (happy, angry, sad) served as three *Deviant* stimuli (each presented with 16.7% probability, 100 trials). For the 300 *Standard* trials, all the 200 words in neutral voice were used, and 100 words were randomly selected for repetition. For each type of the 100 *Deviant* trials, 100 words in each emotional voice were randomly selected and used. The three types of *Deviant* (three emotions) were pseudo-randomly interspersed in between the *standard* trials, with no consecutive *Deviants* trials in the same emotional prosody (see Figure 1 for an example of the sound presentation order). The inter-stimulus interval (ISI) was randomized between 800 – 900 ms, and the total recording time was around 25 minutes.

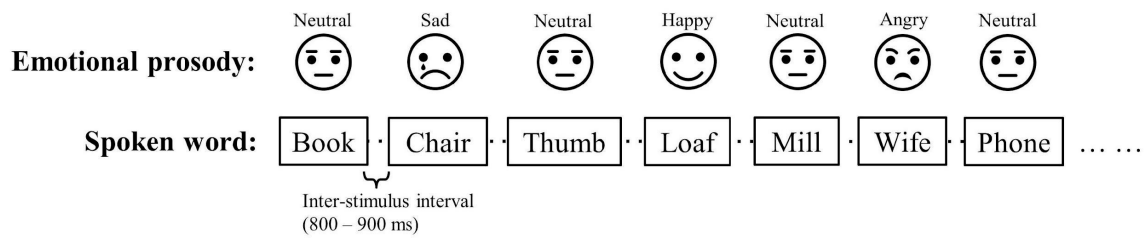


Figure 1. A schematic example of the order of the trials. The *Standard* (neutral prosody) and *Deviant* (angry, happy, and sad prosodies) were always alternating, and the three emotions (*Deviants*) were pseudo-randomly interspersed.

Infants were seated in their parents' laps in an electrically and acoustically treated booth (ETS-Lindgren Acoustic Systems) with a 64-channel WaveGuard EEG cap. One research assistant stayed in the booth and played with silent toys to entertain the infants. A television displaying silent cartoons was also on to keep the infants engaged and still. Parents were instructed to ignore the speech sounds and soothe their children during the EEG recording session. The speech sounds were played via two loudspeakers (M-audio BX8a) placed at a 45-degree azimuth angle, three feet away from the participants. The sound presentation level was set at 65 dBA at the subject's head position (Zhang et al., 2011). The sound presentation was controlled by E-Prime (Psychological Software Tools, Inc) using a Dell PC outside the sound-treated room. Continuous EEG data were recorded through the Advanced Neuro Technology EEG System (Advanced Source Analysis version 4.7). The WaveGuard EEG cap has a layout of 64 Ag/AgCl electrodes following the standard International 10-20 Montage system with intermediate locations, and it is connected to an REFA-72 amplifier (Twente Medical Systems. International B.V. Netherlands). The default bandpass filter for raw data recording was set between 0.016 Hz to 200 Hz, and the sampling rate was 512 Hz. The electrode AFz served as the ground electrode. The impedance of all electrodes was kept under 10 k Ω .

Data analysis

EEG data preprocessing was conducted using EEGLAB v14.1.1 (Delorme & Makeig, 2004). The data underwent low-pass filtering at 40 Hz, downsampling to 250 Hz, and high-pass filtering at 0.5 Hz. Re-referencing was performed to the average of the two mastoid electrodes. The "Clean Rawdata" EEGLAB plug-in was applied to remove low-frequency drifts and non-brain activities, such as muscle activity and sensor motion. Independent Component Analysis (ICA) algorithm (Dammers et al., 2008; Delorme et al., 2001) was then used to attenuate artifacts from eye blinks and other sources. ERP epochs were extracted from 100 ms before stimulus onset to 1000 ms after stimulus onset, and baseline correction was applied using the mean voltages of the 100-ms baseline period. Epochs with data points exceeding ± 150.0 μ V were rejected. ERPLAB v7.0.0 (Lopez-Calderon & Luck, 2014) was used to derive event-related potentials (ERPs) for Standard and each type of Deviant. Difference waveforms were created by subtracting the Standard ERP from each Deviant ERP, resulting in happy, angry, and sad difference waveforms. Infants with less than 30 trials in any condition were excluded from analysis (male = 5, female = 5; mean age = 7.1 months or 218 days).

Statistical analyses were performed using R (<https://www.r-project.org/>) with the "lme4" (Bates et al., 2015) and "lmerTest" (Kuznetsova et al., 2017) packages. Two target components, early mismatch response (early MMR, 100–200 ms) and late mismatch response (late MMR, 300–500 ms), were assessed using the difference waveforms. The time windows were determined based on previous studies and visual inspection. Early and late MMR amplitudes were calculated using the mean voltages within the target windows (early MMN, 100–200 ms; late MMR, 300–500 ms) for statistical analysis. The amplitudes were calculated for channels at frontal (F-line, F3, Fz, F4), central (C-line, C3, Cz, C4), and parietal (P-line, P3, Pz, P4) regions (Kostilainen et al., 2018). Two linear mixed-effect models were performed separately on early and late MMR amplitudes. Each model included by-participant intercept as a random-effect factor and deviant/emotion as a random slope to account for within-infant variations. Deviant/Emotion (happy, sad, and angry), region of the

electrode (anterior, central, and parietal), and laterality of the electrode (left, middle, and right) were included as trial-level fixed-effect factors. Infants' biological sex (female and male) and age (continuous variable in months) were included as participant-level fixed factors. Finally, cross-level interactions of emotion and sex and emotion and age were also included.

Results

To facilitate visualization, infants were divided into younger and older age groups using median-split, but age was treated as a continuous variable in the statistical analyses. Figure 2 displays the average difference waveforms representing early and late MMRs to emotional prosody changes. Figure 3 presents the ERPs before the subtraction process. Additionally, Figures 2 and 3 show the difference waveforms and grand mean ERP waveforms recorded from frontal (F-line: F3, Fz, F4), central (C-line: C3, Cz, C4), and parietal (P-line: P3, Pz, P4) electrodes for all emotional prosodies. The topographic maps of early and late MMRs for each emotional prosody are depicted in Figure 4.

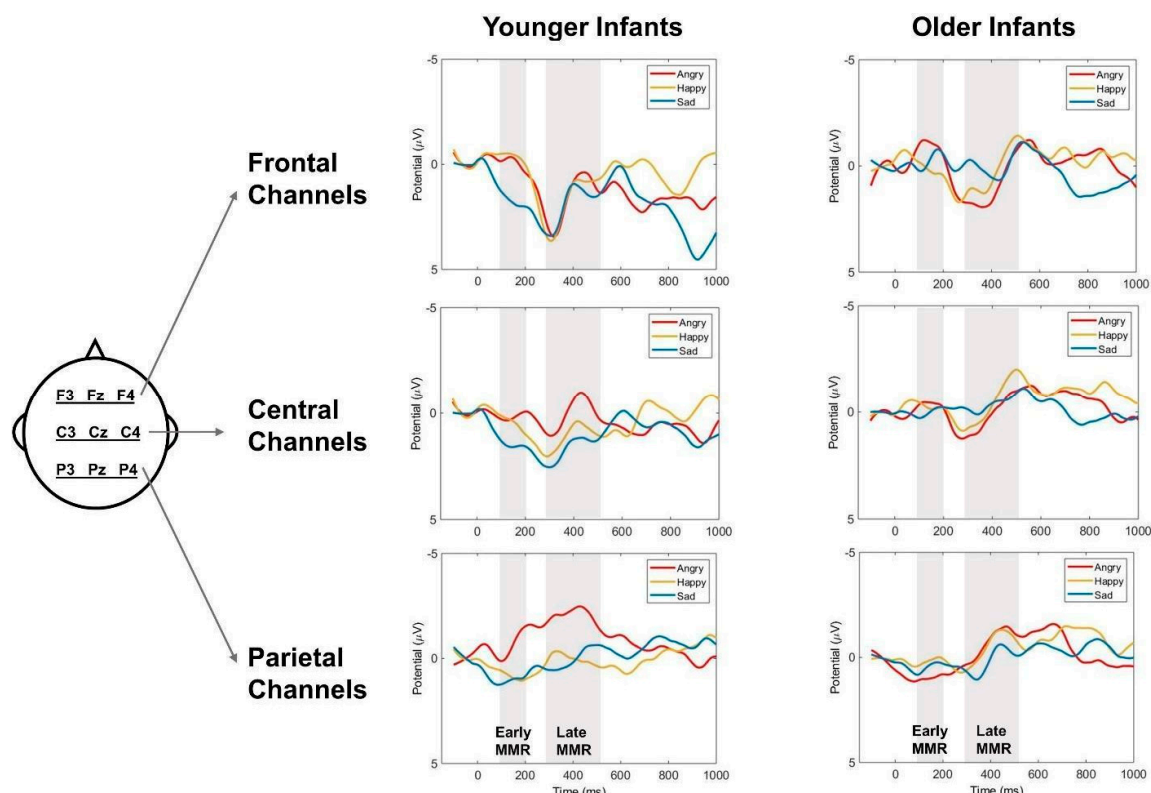


Figure 2. The grand mean difference waveforms of angry, happy, and sad in younger and older infant listeners (split by median age at 8.2-month). Mean amplitudes of the F-line (F3, Fz, F4), C-line (C3, Cz, C4), P-line (P3, Pz, P4) electrodes were used for the waveforms. The gray shaded areas mark the windows for early mismatch response (early MMR, 100 – 200 ms) and late MMR (300 – 500 ms).

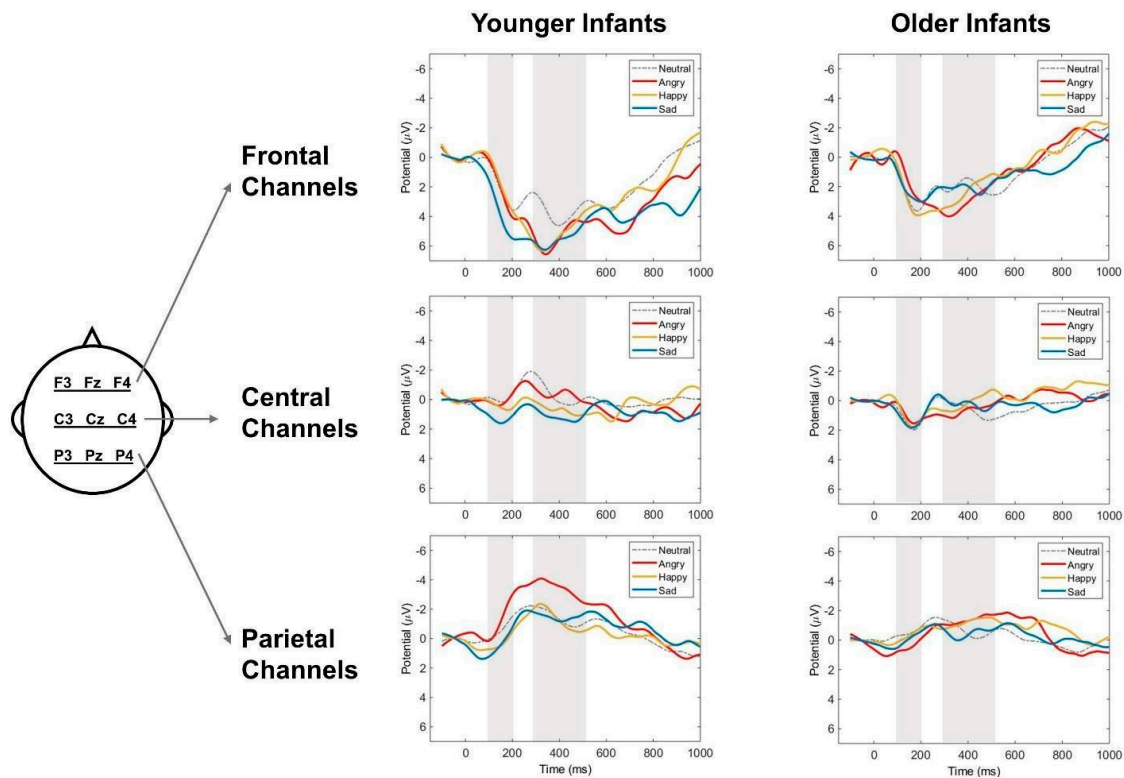


Figure 3. The grand mean event-related potential (ERP) waveforms of *Standard* (neutral prosody) and *Deviant*s (angry, happy, and sad) in younger and older infant listeners. Mean amplitudes of the F-line (F3, Fz, F4), C-line (C3, Cz, C4), P-line (P3, Pz, P4) electrodes were used for the waveforms. The gray shaded areas mark the windows for early mismatch response (early MMR, 100 – 200 ms) and late MMR (300 – 500 ms).

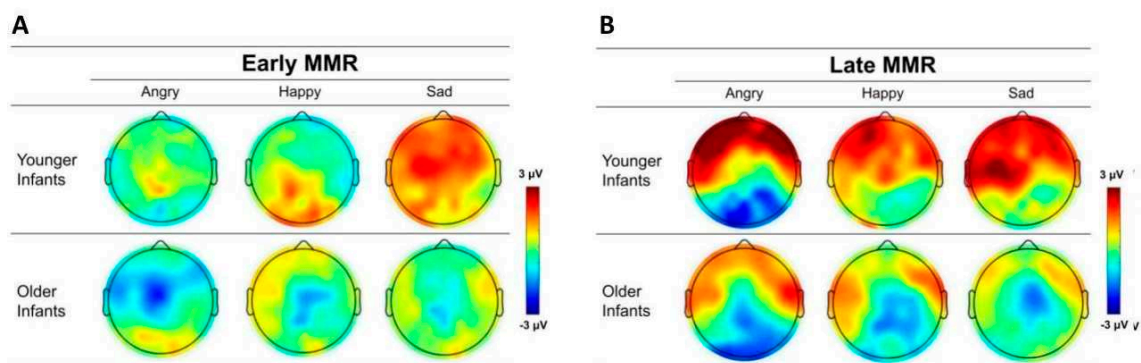


Figure 4. The scalp topographic maps of (A) early mismatch response (MMR) and (B) late MMR to angry, happy, and sad emotional prosodies averaged across younger and older infant listeners. The topographies are based on the average values in each component window (early MMR, 100 – 200 ms; late MMR, 300 – 500 ms).

Early Mismatch Response (Early MMR)

The main effect of emotion showed that early MMRs were significantly more negative to happy ($\beta = -4.79$, $t = -2.04$, $p = 0.05$) and to angry ($\beta = -5.62$, $t = -2.97$, $p = 0.005$) speech compared with sad speech. The electrode laterality effect showed that early MMRs were significantly more negative at midline than the left electrodes ($\beta = 0.46$, $t = 1.93$, $p = 0.05$). The age effect was significant, showing that infants' early MMRs were more negative as their age increased ($\beta = -0.41$, $t = -2.11$, $p = 0.04$). The sex effect also revealed that male infants showed significantly more negative early MMRs compared with female infants ($\beta = -1.61$, $t = -2$, $p = 0.05$). Furthermore, the interaction between age and emotion

showed distinct early MMRs between angry and sad prosody in younger infants, but the difference became more subtle in older infants ($\beta = 0.53$, $t = 2.35$, $p = 0.02$). The model of early MMR amplitudes is summarized in Table 2 (See supplementary material). Figure 5 shows the main effects of emotion, electrode laterality, age, sex, and the interaction between emotion and age.

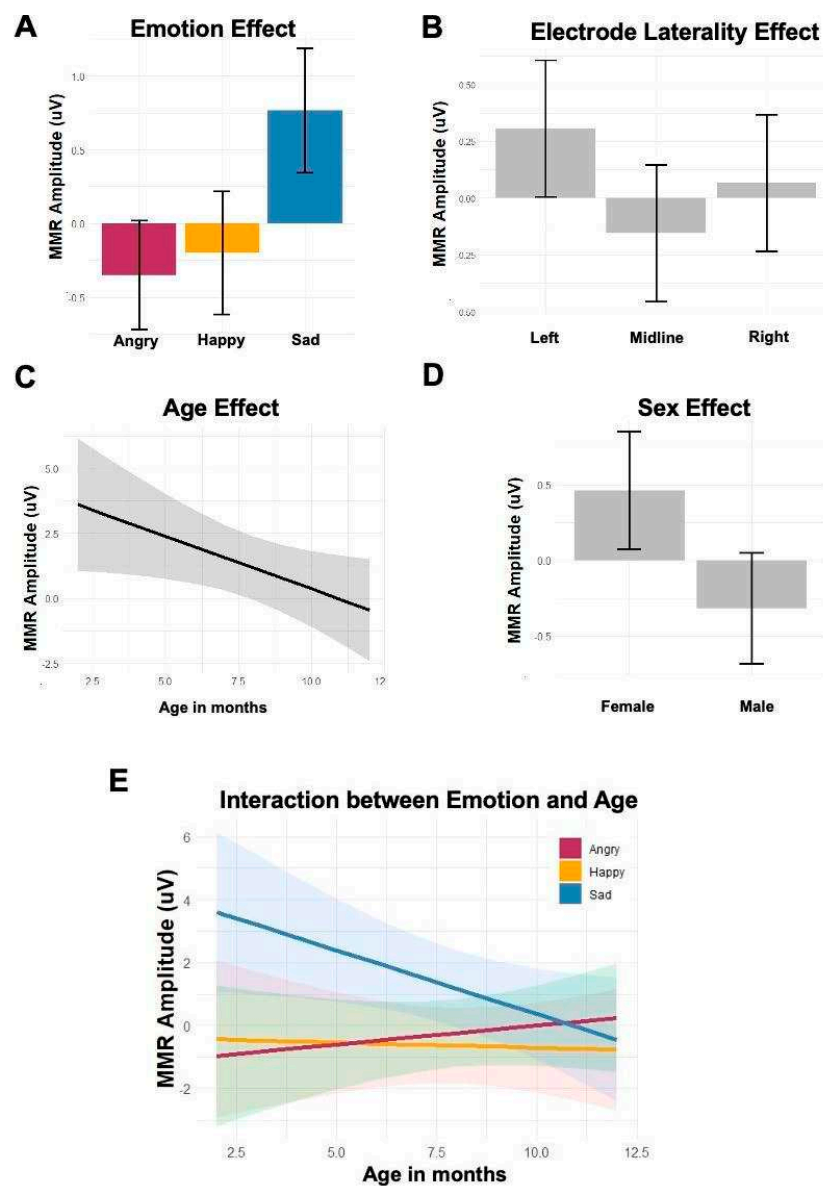


Figure 5. The main effects of (A) Emotion, (B) Electrode Laterality, (C) Age, (D) Sex, and (E) the interaction between Emotion and Age in the linear-mixed effect model with early MMR amplitudes.

Late Mismatch Response (Late MMR)

The main effect of emotion showed that late MMRs were significantly more negative to angry speech ($\beta = -6.16$, $t = -2.55$, $p = 0.02$) compared with sad speech. The electrode region effect showed that late MMRs were more negative toward the posterior region compared to the central region ($\beta = -1.03$, $t = -3.61$, $p < 0.001$), and more negative toward the central region compared with the frontal region ($\beta = 1.30$, $t = 4.56$, $p < 0.001$). The electrode laterality effect showed that late MMRs were significantly more negative at midline than the left electrodes ($\beta = 0.79$, $t = 2.79$, $p = 0.005$). Neither age nor sex effect was significant. The interaction between age and emotion showed distinct late MMRs between angry and sad prosody in younger infants, but the difference became more subtle in older infants ($\beta = 0.60$, $t = 2.12$, $p = 0.01$). The model of late MMR amplitudes is summarized in Table 3 (See

supplementary material). Figure 6 shows the main effects of emotion, electrode region, electrode laterality, and the interaction between emotion and age.

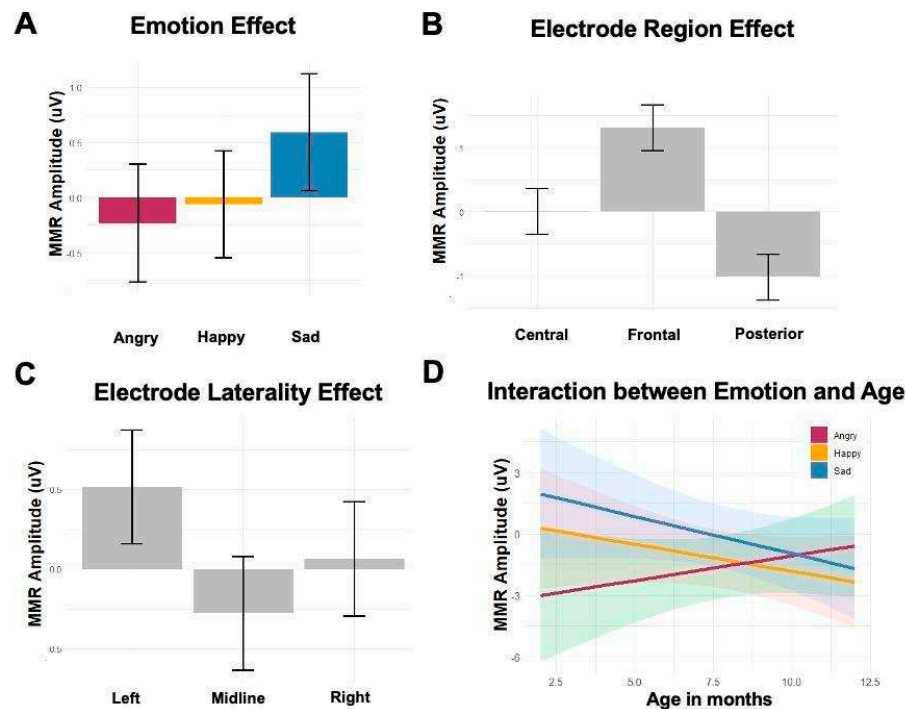


Figure 6. The main effects of (A) Emotion, (B) Electrode Region, (C) Electrode Laterality, and (D) the interaction between Emotion and Age in the linear-mixed effect model with late MMR amplitudes.

Discussion

The current study employed a modified multi-feature oddball paradigm to investigate 3-to-11-month-old infants' neural sensitivities to angry, happy, and sad prosodies against neutral prosody over non-repeating English spoken words. The findings confirmed our hypothesis that infants at this age can perceptually group emotional categories and automatically detect prosodic changes across different words, indicating early processing of emotional information independent of attention. These results provide evidence for the elicitation of mismatch responses based on emotional prosodic regularities, demonstrating the influence of abstract auditory features on neural processing (Näätänen et al., 2001).

Early MMR (100 – 200 ms)

The early Mismatch Negativity Response (MMR) reflects the neural matching process of comparing incoming sounds with those stored in auditory memory (Näätänen et al., 2007). In early infancy, it starts as a positive deflection and gradually becomes more negative (Csibra et al., 2008; Kushnerenko et al., 2013). Our findings showed that sad prosody elicited a more positive early MMR compared to angry and happy voices, supporting the notion that infants older than 3 months can pre-attentively differentiate sad voices from neutral ones. Similar early MMRs to angry and happy prosody changes suggest that these high-arousal emotions may be processed similarly at this early stage of sensory processing. These distinct early MMRs to emotional speech changes confirm that infants between 3 and 11 months of age can pre-attentively detect acoustic and abstract vocal categories, unlike the indistinguishable emotional MMRs observed in newborns (Kostilainen et al., 2018). The developmental changes in MMR were further supported by the age effect, where older infants exhibited more negative MMR amplitudes compared to younger infants.

We observed an interaction between emotion and age in the early MMR, showing a more positive deflection to sad voices than angry voices in younger infants, but this difference became less pronounced as infants aged. In other words, older infants exhibited similar MMRs to all three vocal

emotions. Distinguishing between happy and angry voices was not observed, which partially explains the confusion experienced by young infants when listening to these two high-arousal emotions (Flom & Bahrick, 2007; Walker-Andrews & Lennon, 1991). The differential age effect on angry and sad voices suggests changes in infants' neural sensitivities over time. Our data demonstrate that infants under one year old are already capable of distinguishing various emotions in human voice at an early neural processing stage, representing an improvement from the undifferentiated early MMR observed in newborns (Kostilainen et al., 2018). Furthermore, the less distinct neural representations of different vocal emotions with age may indicate that older infants, with more listening experience, detect affective prosody changes in a more generalized manner. It is also plausible that the complexity of changing spoken word stimuli involves more age-related linguistic processing that attenuates the discriminative responses to emotional prosody patterns.

We observed a sex effect on infants' early MMRs, which was not found in a previous behavioral study using preferential looking measures (Kao & Zhang, 2022). Contrary to our hypothesis, male infants exhibited more negative early MMR amplitudes (-0.32 uV, $SE = 0.37$ uV) compared to female infants (0.46 uV, $SE = 0.39$ uV). A previous study on newborns found emotion-specific sex differences in early MMRs, with males showing more negative MMRs to angry and sad emotions than females (Kostilainen et al., 2018). Another study on preterm neonates reported that female infants had stronger MMRs to voice changes, including emotional voices, than males (Kostilainen et al., 2021). However, since infant MMRs can manifest as either positive or negative deflections, caution is needed in interpreting the sex differences observed in the current study with infants aged 3 to 12 months (Etchell et al., 2018). The early MMR primarily reflects lower-level sensory processing, and further research with larger sample sizes is needed to determine whether the sex effect on this neurophysiological component implies functional differences in automatic auditory processing of emotional prosody between male and female infants (McClure, 2000; Wallentin, 2009). There was no interaction between sex and emotional categories, suggesting that the differential early MMR amplitudes in male and female infants were not specific to certain emotions.

To the best of our knowledge, this is the first study using a multi-feature oddball design to examine emotion-modulated MMRs in early infancy. Previous research with a similar design did not find an emotional effect on early MMRs in newborns, even though vocal emotions conveyed by simple syllables "ta-ta" should be more easily discriminable (Kostilainen et al., 2018). Our data demonstrate that infants under one year old already exhibit neural sensitivities to acoustic pattern grouping for vocal emotions. Moreover, these automatic voice processing abilities are manifested in distinct early MMR amplitudes for different emotions, with more negative deflections for angry and happy prosodies compared to sad prosody. The distinct early MMR responses to different vocal emotions indicate the rapid development of infants' auditory system, enabling efficient processing of complex vocal emotions with only a few months of listening experience.

Late MMR (300 – 500 ms)

The late MMR in infants is often considered as their version of MMN (Cheour et al., 2000; Leppänen et al., 2004; Trainor, 2010). Unlike the adult MMN, infants' MMR emerges as a slow positive wave and gradually develops into a negative deflection over time (He et al., 2009). Previous studies on pre-attentive neural responses to emotional prosody in infants have mainly focused on this late MMR (300 - 500 ms) rather than the early MMR (100 - 200 ms) (Cheng et al., 2012; Kostilainen et al., 2020; Kostilainen et al., 2018; Zhang et al., 2014). It has been suggested that this slow positive wave may represent a combination of MMR and a subsequent component known as P3a, which is a fronto-central positive wave elicited by novel events in adults (Escera et al., 2000; Friedman et al., 2001). The current data confirmed the presence of this fronto-centrally oriented late MMR with a positive deflection, indicating that the multi-feature oddball task successfully elicited the targeted MMR, as previously observed in adult listeners (Kao & Zhang, 2023). We found that our infant listeners exhibited a stronger and more negative-oriented MMR to angry voices compared to sad voices, particularly among younger infants (see Figures 2 and 6C). This finding aligns with previous studies conducted with newborns by Cheng et al. (2012) and Zhang et al. (2014). One possible explanation is

that infants' auditory systems are naturally attuned to respond more efficiently to negative and threat-related signals, which is reflected in their stronger late MMR to angry voices. Notably, our study utilized varying spoken words, rather than simple syllables, to convey emotional prosody, which increased the demands for prosodic pattern extraction based on statistical regularities in the acoustic parameters. Despite this added challenge, infants in our study successfully extracted the relevant emotional prosodic categories across changing words and exhibited distinct neural responses in their pre-attentive system.

In the late MMR window, we observed an interaction between emotion and age. The late MMRs to angry prosody were significantly more negative than the MMRs to sad prosody, particularly in younger infants. As infants' age increased, the negative deflection of MMRs to angry prosody approached zero, and MMRs to different emotional changes became similar in older infants. To our knowledge, there have been no previous reports on older infants' MMRs to vocal emotions (happy, angry, and sad) using a multi-feature oddball task. The closest study by Kostilainen et al. (2020) on newborns found similar MMR amplitudes to angry and sad voices when emotional vocal events were rare in the experimental paradigm, which contradicts our findings of distinguishable late MMRs to angry and sad prosody in younger infants with more listening experience than neonates. Our data suggest that younger infants are most sensitive to angry prosody, while older infants process different emotional changes similarly without exhibiting distinguishable late MMR amplitudes. Another study by Grossmann et al. (2005) measured seven-month-old infants' ERPs to randomly presented angry and happy voices and found a more negative ERP response to angry prosody, consistent with our findings (Figure 6D). Although our experimental protocol differed from that of Grossmann et al. (2005), it is noteworthy that they also used multiple words (up to 74) to convey each vocal emotion. Taken together, we are confident that it is feasible to use a more complex and natural linguistic context to investigate infants' neural sensitivities to emotional prosody. Future studies examining MMRs to emotional voices in older infants (7-12 months old) are encouraged to further examine our current findings.

Unlike the early MMR, we did not observe a biological sex effect on infants' late MMRs to emotional voices. Previous studies on newborns' MMR in this later window to vocal emotional changes either did not find sex differences (Cheng et al., 2012; Kostilainen et al., 2018) or did not include sex as a factor to explain variations in infants' MMR amplitudes (Grossmann et al., 2005; Kostilainen et al., 2020; Zhang et al., 2014). The similar vocal emotional MMRs in male and female newborns do not provide clear evidence of biological sex differences in emotional voice processing during the first year of life.

Limitations and Future Directions

The current study examines infants' neural sensitivities to emotional prosodies presented with non-repeating spoken words. While our findings indicate that infants can differentiate among the four basic emotional prosodies within varying linguistic contexts, it is important to note that our results do not provide insight into whether these prosodic categories correspond to infants' subjective emotional experiences as perceived and interpreted by adults. This limitation is common in infant MMR studies, as the paradigm used only requires participants to automatically detect different voices without active engagement. To further explore the potential involvement of emotional evaluation, future research could incorporate audiovisual emotional stimuli to assess infants' cross-modal emotional congruency detection and record their EEG signals (Flom & Whiteley, 2014; Grossman, 2013; Otte et al., 2015). Although including both behavioral and EEG tasks adds complexity to the experimental design, it allows for the measurement of neurophysiological responses underlying infants' emotional evaluation.

Another limitation of our study is that the speech stimuli were obtained from a single female speaker, which restricts the generalizability of the findings to real-life scenarios where infants encounter emotional speech from multiple speakers. However, the rationale behind using non-repeating spoken words from a single speaker in our study was to introduce a more diverse yet manageable linguistic context for infants to extract emotional prosody. Previous studies on emotional

multi-feature oddball tasks for infants typically utilized a small set of simple and fixed syllables (Cheng et al., 2012; Kostilainen et al., 2020; Kostilainen et al., 2018; Zhang et al., 2014), making it reasonable to first examine infants' emotional MMRs to varying words from the same speaker before expanding to multiple speakers.

Despite these limitations, our findings, along with the report by Grossmann and colleagues (2005), demonstrate that infants' pre-attentive neural systems are capable of perceptually grouping words with the same emotional prosody in contrast to those conveying a different emotional prosody. This is a crucial indicator of how infants may utilize the prosodic dynamics of language input for their social and communicative development. Future studies can incorporate speech stimuli from male speakers to create an even more natural listening context and thoroughly investigate the impact of biological sex on infants' early processing of emotional voices. Additionally, further exploration is needed to understand how infants' neural sensitivity to prosodic emotion processing contributes to their subsequent emotional development, extraction of affective meaning, vocabulary acquisition, and other aspects of socio-linguistic development (Lindquist & Gendron, 2013; Morton & Trehub, 2003; Nencheva et al., 2023; Quam & Swingley, 2012).

Feasibility of the efficient passive-listening multi-feature oddball paradigm for infants has great implications for future studies investigating the role of speech prosody including emotional voice processing in the development of language and speech communication in populations with short sustained attention to lengthy tasks (Ding & Zhang, 2023). For instance, Korpilahti et al. (2007) studied the differences in neural responses to angry voices between children with and without Asperger syndrome. By incorporating the multi-feature design, researchers can examine more emotional voices and understand their different effects on speech processing and socioemotional development in neurodivergent infants and children (Zhang et al., 2022).

Conclusion

The current study demonstrates the feasibility of the multi-feature oddball paradigm in studying early perception of emotional prosody in speech. By utilizing non-repeating spoken words, we successfully elicited both early and late MMRs in infants in response to happy, angry, and sad prosodies. The findings clearly indicate distinct developmental changes in infants' neural responses to each emotional category, highlighting the sensitivity of EEG as a tool for capturing developmental trends that may not be evident in behavioral studies. Additionally, we observed differences in MMR amplitudes between male and female infants in the early MMR window, but not in the late MMR window. Further research is necessary to elucidate the role of biological and social factors in the observed sex differences in processing socio-emotional signals as well as the functional implications of the interaction between age and emotion category effects in the early and late MMRs for language learning and socio-emotional development.

Funding information: University of Minnesota's Graduate Research Partnering Project Award, Graduate Dissertation Fellowship, Brain Imaging Grant and SEED Grant, and the Bryng Bryngelson Fund from the Department of Speech-Language-Hearing Sciences.

DATA AVAILABILITY STATEMENT: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements: This work received financial support from the University of Minnesota, including the Graduate Research Partnering Project (CK), Graduate Dissertation Fellowship (CK), and Brain Imaging Grant from the College of Liberal Arts (YZ), and the Bryng Bryngelson Fund from the Department of Speech-Language-Hearing Sciences (CK). We would like to thank Dr. Maria Sera for help with subject recruitment via Institute of Child Development's infant pool and Jessica Tichy, Natasha Stark, Shannon Hofer-Pottala, Kailie McGuigan, Emily Krattley, Hayley Levenhagen, and Megan Peterson for assistance with the data collection.

CONFLICTS OF INTEREST STATEMENT: The authors declare no conflict of interests.

ETHICS STATEMENT: The study was conducted with written informed consent obtained from legal guardians in compliance with the Institutional Review Board approval at the University of Minnesota.

References

- Amorim, M., Anikin, A., Mendes, A. J., Lima, C. F., Kotz, S. A., & Pinheiro, A. P. (2019). Changes in vocal emotion recognition across the life span. *Emotion*.
- Aslin, R. N. (2007). What's in a look? *Developmental Science*, 10(1), 48-53.
- Banase, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, 12(1), 2.
- Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G. J., Renvall, V., & Deoni, S. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, 21(14), 1220-1224.
- Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer [Computer program]*. Version 6.1.09.
- Cheng, Y., Lee, S.-Y., Chen, H.-Y., Wang, P.-Y., & Decety, J. (2012). Voice and emotion processing in the human neonatal brain. *Journal of Cognitive Neuroscience*, 24(6), 1411-1419.
- Cheour, M. (2007). Development of mismatch negativity (MMN) during infancy. In M. de Haan (Ed.), *Infant EEG and event-related potentials* (pp. 171-198). Psychology Press.
- Cheour, M., Kushnerenko, E., Ceponiene, R., Fellman, V., & Näätänen, R. (2002). Electric brain responses obtained from newborn infants to changes in duration in complex harmonic tones. *Developmental neuropsychology*, 22(2), 471-479.
- Cheour, M., Leppänen, P. H., & Kraus, N. (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clinical Neurophysiology*, 111(1), 4-16.
- Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of learning and memory*, 92(2), 225-234.
- Csibra, G., Kushnerenko, E., & Grossmann, T. (2008). Electrophysiological methods in studying infant cognitive development.
- D'Entremont, B., & Muir, D. (1999). Infant responses to adult happy and sad vocal and facial expressions during face-to-face interactions. *Infant Behavior and Development*, 22(4), 527-539.
- Dammers, J., Schiek, M., Boers, F., Silex, C., Zvyagintsev, M., Pietrzyk, U., & Mathiak, K. (2008). Integration of amplitude and phase statistics for complete artifact removal in independent components of neuromagnetic recordings. *IEEE transactions on biomedical engineering*, 55(10), 2353-2362.
- De Haan, M. (2002). Introduction to infant EEG and event-related potentials. *Infant EEG and event-related potentials*, 39-76.
- De Haan, M. (Ed.). (2013). *Infant EEG and event-related potentials*. Psychology Press.
- Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience*, 12(3), 449-460.
- Delorme, A., & Makeig, S. (2004, 2004/03/15/). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21. <https://doi.org/https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Delorme, A., Makeig, S. & Sejnowski, T. (2001). Automatic artifact rejection for EEG data using high-order sStatistics and independent component analysis. *Proceedings of the 3rd International Independent Component Analysis and Blind Source Decomposition Conference*, December 9-12, San Diego (USA).
- Ding, H., & Zhang, Y. (2023). Speech prosody in mental disorders. *Annual Review of Linguistics*, 9, 335-355. <https://doi.org/10.1146/annurev-linguistics-030421-065139>
- Dupuis, K., & Pichora-Fuller, M. K. (2010). *Toronto Emotional Speech Set (TESS)*. University of Toronto, Psychology Department. <https://doi.org/10.5683/SP2/E8H2MF>
- Etchell, A., Adhikari, A., Weinberg, L. S., Choo, A. L., Garnett, E. O., Chow, H. M., & Chang, S. E. (2018). A systematic literature review of sex differences in childhood language and brain development. *Neuropsychologia*, 114, 19-31.
- Escera, C., Alho, K., Schröger, E., & Winkler, I. W. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neurotology*, 5(3-4), 151-166.
- Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497-1510.
- Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, 64(3), 657-674.

- Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, 43(1), 238. <https://doi.org/10.1037/0012-1649.43.1.238>
- Flom, R., & Whiteley, M. O. (2014, Jan). The dynamics of intermodal matching: Seven- and 12-month-olds' intermodal matching of affect. *European Journal of Developmental Psychology*, 11(1), 111-119.
- Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *Neuroreport*, 13(10), 1251-1254.
- Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001, 2001/06/01/). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, 25(4), 355-373. [https://doi.org/https://doi.org/10.1016/S0149-7634\(01\)00019-7](https://doi.org/https://doi.org/10.1016/S0149-7634(01)00019-7)
- García-Sierra, A., Ramírez-Esparza, N., Wig, N., & Robertson, D. (2021). Language learning as a function of infant directed speech (IDS) in Spanish: Testing neural commitment using the positive-MMR. *Brain and Language*, 212, 104890.
- Grossman, T. (2013). The early development of processing emotions in face and voice. In *Integrating face and voice in person perception* (pp. 95-116). Springer Science + Business Media; US.
- Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, 28(2), 219-236. <https://doi.org/10.3233/RNN-2010-0499>
- Grossmann, T., Striano, T., & Friederici, A. D. (2005, Nov). Infants' electric brain responses to emotional prosody. *NeuroReport: For Rapid Communication of Neuroscience Research*, 16(16), 1825-1828. <https://doi.org/10.1097/01.wnr.0000185964.34336.b1>
- Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science*, 9(3), 309-315. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-7687.2006.00494.x>
- Groves, P. M., & Thompson, R. F. (1970). Habituation: a dual-process theory. *Psychological review*, 77(5), 419.
- Hartkopf, J., Moser, J., Schleger, F., Preissl, H., & Keune, J. (2019). Changes in event-related brain responses and habituation during child development—A systematic literature review. *Clinical Neurophysiology*, 130(12), 2238-2254.
- He, C., Hotson, L., & Trainor, L. J. (2009). Maturation of cortical mismatch responses to occasional pitch change in early infancy: effects of presentation rate and magnitude of change. *Neuropsychologia*, 47(1), 218-229.
- Hohenberger, A. (2011). The role of affect and emotion in language development. In *Affective computing and interaction: Psychological, cognitive and neuroscientific perspectives* (pp. 208-243). IGI Global.
- Hung, A.-Y., & Cheng, Y. (2014). Sex differences in preattentive perception of emotional voices and acoustic attributes. *Neuroreport*, 25(7), 464-469.
- Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. *Handbook of emotions*, 2, 220-235.
- Kao, C., Seva, M., & Zhang, Y. (2022). Emotional speech processing in 3- to 12- month old infants: Influences of emotion categories and acoustic parameters. *Journal of Speech, Language and Hearing Research*, 65, 487-500. https://doi.org/10.1044/2021_JSLHR-21-00234
- Kao, C., & Zhang, Y. (2023). Detecting emotional prosody in real words: Electrophysiological evidence from a modified multi-feature oddball paradigm. *Journal of Speech, Language and Hearing Research*. https://doi.org/10.1044/2023_JSLHR-22-00652
- Kok, T. B., Post, W. J., Tucha, O., de Bont, E. S., Kamps, W. A., & Kingma, A. (2014). Social competence in children with brain disorders: A meta-analytic review. *Neuropsychology review*, 24(2), 219-235.
- Korpilahti, P., Jansson-Verkasalo, E., Mattila, M.-L., Kuusikko, S., Suominen, K., Rytty, S., Pauls, D. L., & Moilanen, I. (2007). Processing of affective speech prosody is impaired in Asperger syndrome. *Journal of autism and developmental disorders*, 37(8), 1539-1549.
- Kostilainen, K., Partanen, E., Mikkola, K., Wikström, V., Pakarinen, S., Fellman, V., & Huotilainen, M. (2020). Neural processing of changes in phonetic and emotional speech sounds and tones in preterm infants at term age. *International Journal of Psychophysiology*, 148, 111-118.
- Kostilainen, K., Partanen, E., Mikkola, K., Wikström, V., Pakarinen, S., Fellman, V., & Huotilainen, M. (2021). Repeated parental singing during kangaroo care improved neural processing of speech sound changes in preterm infants at term age. *Frontiers in Neuroscience*, 15, 686027.
- Kostilainen, K., Wikström, V., Pakarinen, S., Videman, M., Karlsson, L., Keskinen, M., Scheinin, N. M., Karlsson, H., & Huotilainen, M. (2018). Healthy full-term infants' brain responses to emotionally and linguistically

- relevant sounds using a multi-feature mismatch negativity (MMN) paradigm. *Neuroscience letters*, 670, 110–115. <https://doi.org/10.1016/j.neulet.2018.01.039>
- Kushnerenko, E. V., Van den Bergh, B. R., & Winkler, I. (2013). Separating acoustic deviance from novelty during the first year of life: a review of event-related potential evidence. *Frontiers in Psychology*, 4, 595.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>
- Lawrence, L., & Fernald, A. (1993). When prosody and semantics conflict: Infants' sensitivity to discrepancies between tone of voice and verbal content. Poster presented at the Biennial Meeting of the Society for Research in Child Development,
- Leppänen, P. H., Guttorm, T. K., Pihko, E., Takkinen, S., Eklund, K. M., & Lyytinen, H. (2004). Maturation effects on newborn ERPs measured in the mismatch negativity paradigm. *Experimental Neurology*, 190, 91–101.
- Leppänen, J. M., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child development*, 78(1), 232–245. <https://doi.org/10.1111/j.1467-8624.2007.00994.x>
- Lindquist, K. A., & Gendron, M. (2013). What's in a Word? Language Constructs Emotion Perception. *Emotion Review*, 5(1), 66–71. <https://doi.org/10.1177/1754073912451351>
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213.
- Mani, N., & Pätzold, W. (2016). Sixteen-month-old infants' segment words from infant-and adult-directed speech. *Language Learning and Development*, 12(4), 499–508. <https://doi.org/10.1080/15475441.2016.1171717>
- Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 35(3), 204–214. [https://doi.org/10.1002/\(SICI\)1098-2302\(199911\)35:3<204::AID-DEV5>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1098-2302(199911)35:3<204::AID-DEV5>3.0.CO;2-V)
- Maurer, U., Bucher, K., Brem, S., & Brandeis, D. (2003). Development of the automatic mismatch response: from frontal positivity in kindergarten children to the mismatch negativity. *Clinical Neurophysiology*, 114(5), 808–817.
- McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, 126(3), 424.
- Mehler, J., Bertoncini, J., Barriere, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, 7(5), 491–497.
- Minagawa-Kawai, Y., Van Der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebral Cortex*, 21(2), 254–261.
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16(4), 495–500.
- Morningstar, M., Nelson, E. E., & Dirks, M. A. (2018). Maturation of vocal emotion recognition: Insights from the developmental and neuroimaging literature. *Neuroscience & Biobehavioral Reviews*, 90, 221–230.
- Morton, J.B., & Trehub, S.E. (2001). Children's Understanding of Emotion in Speech. *Child Development*, 72, 834–843. <https://doi.org/10.1111/1467-8624.00318>
- Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development*, 67(6), 3219–3237.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, 118(12), 2544–2590.
- Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology*, 115(1), 140–144.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, 24(5), 283–288.
- Nelson, C. A., & De Haan, M. (1996). Neural correlates of infants' visual responsiveness to facial expressions of emotion. *Developmental psychobiology*, 29(7), 577–595. [https://doi.org/10.1002/\(SICI\)1098-2302\(199611\)29:7<577::AID-DEV3>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1098-2302(199611)29:7<577::AID-DEV3>3.0.CO;2-R)
- Nencheva, M. L., Tamir, D. I., & Lew-Williams, C. (2023). Caregiver speech predicts the emergence of children's emotion vocabulary. *Child Development*, 94, 585–602. <https://doi.org/10.1111/cdev.13897>

- Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, 11(3), 255-268.
- Otte, R. A., Donkers, F. C. L., Braeken, M. A. K. A., & Van den Bergh, B. R. H. (2015, 2015/04/01/). Multimodal processing of emotional information in 9-month-old infants I: Emotional faces and voices. *Brain and Cognition*, 95, 99-106. <https://doi.org/https://doi.org/10.1016/j.bandc.2014.09.007>
- Quam, C., & Swingle, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child development*, 83(1), 236–250. <https://doi.org/10.1111/j.1467-8624.2011.01700.x>
- Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport*, 16(6), 635-639.
- Shafer, V. L., Yan, H. Y., & Garrido-Nag, K. (2012). Neural mismatch indices of vowel discrimination in monolingually and bilingually exposed infants: Does attention matter? *Neuroscience Letters*, 526(1), 10-14.
- Soken, N. H., & Pick, A. D. (1992). Intermodal Perception of Happy and Angry Expressive Behaviors by Seven-Month-Old Infants. *Child Development*, 63(4), 787-795.
- Soken, N. H., & Pick, A. D. (1999). Infants' perception of dynamic affective expressions: do infants distinguish specific expressions? *Child Development*, 70(6), 1275-1282.
- Thönnessen, H., Boers, F., Dammers, J., Chen, Y.-H., Norra, C., & Mathiak, K. (2010). Early sensory encoding of affective prosody: neuromagnetic tomography of emotional category changes. *Neuroimage*, 50(1), 250-259.
- Tillman, T. W., & Carhart, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6.*
- Trainor, L. J. (2010). Using electroencephalography (EEG) to measure maturation of auditory cortex in infants: Processing pitch, duration and sound location. Tremblay, RE, Barr, RG, Peters, R. deV., Boivin, M.(Eds.), *Encyclopedia on Early Childhood Development*. Centre of Excellence for Early Childhood Development, Montreal, Quebec, 1-5.
- Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science*, 7(3), 261-269.
- Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology*, 22(3), 373. <https://doi.org/10.1037/0012-1649.22.3.373>
- Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior & Development*. [https://doi.org/10.1016/S0163-6383\(83\)90331-4](https://doi.org/10.1016/S0163-6383(83)90331-4)
- Walker-Andrews, A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior and Development*, 14(2), 131-142. [https://doi.org/10.1016/0163-6383\(91\)90001-9](https://doi.org/10.1016/0163-6383(91)90001-9)
- Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, 33(3), 514-535.
- Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and Language*, 108(3), 175-183.
- Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huottilainen, M., Näätänen, R., & Sussman, E. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences*, 100(20), 11812-11815.
- 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.
- Zhang, D., Liu, Y., Hou, X., Sun, G., Cheng, Y., & Luo, Y. (2014). Discrimination of fearful and angry emotional voices in sleeping human neonates: a study of the mismatch brain responses. *Frontiers in behavioral neuroscience*, 8, 422.
- Zhang, M., Xu, S., Chen, Y., Lin, Y., Ding, H., & Zhang, Y. (2022). Recognition of affective prosody in autism spectrum conditions: A systematic review and meta-analysis. *Autism*, 26(4), 798–813. <https://doi.org/10.1177/1362361321995725>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.