



PAPER

Neural coding of formant-exaggerated speech in the infant brain

Yang Zhang,^{1,2} Tess Koerner,¹ Sharon Miller,¹ Zach Grice-Patil,³ Adam Svec,¹ David Akbari,¹ Liz Tusler¹ and Edward Carney¹

1. Department of Speech-Language-Hearing Sciences, University of Minnesota, USA

2. Center for Neurobehavioral Development, University of Minnesota, USA

3. Department of Psychology, University of Minnesota, USA

Abstract

Speech scientists have long proposed that formant exaggeration in infant-directed speech plays an important role in language acquisition. This event-related potential (ERP) study investigated neural coding of formant-exaggerated speech in 6–12-month-old infants. Two synthetic /i/ vowels were presented in alternating blocks to test the effects of formant exaggeration. ERP waveform analysis showed significantly enhanced N250 for formant exaggeration, which was more prominent in the right hemisphere than the left. Time-frequency analysis indicated increased neural synchronization for processing formant-exaggerated speech in the delta band at frontal-central-parietal electrode sites as well as in the theta band at frontal-central sites. Minimum norm estimates further revealed a bilateral temporal-parietal-frontal neural network in the infant brain sensitive to formant exaggeration. Collectively, these results provide the first evidence that formant expansion in infant-directed speech enhances neural activities for phonetic encoding and language learning.

Introduction

Language input is assigned roles of varying importance in acquisition models and theories. The ‘poverty-of-stimulus’ argument asserts that language is unlearnable from the impoverished input data available to children (Chomsky, 1980). In contrast, speech research over the past five decades has established that enriched exposure adaptively guides language acquisition early in life (Höhle, 2009; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson, 2008). For example, when talking to infants, people across cultures tend to use exaggerated pitch, elongated words, and expanded vowel space with stretched formant frequencies (Ferguson, 1964; Fernald, 1992; Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, Stolyarova, Sundberg & Lacerda, 1997). This special speech style undergoes important age-related changes to accommodate the communicative capacity of the developing mind (Amano, Nakatani & Kondo, 2006; Fernald & Morikawa, 1993; Kitamura, Thanavishuth, Burnham & Luksaneeyanawin, 2001; Liu, Tsao & Kuhl, 2009).

The acoustic alterations of infant-directed speech (IDS) purportedly serve vital social and linguistic func-

tions in early learning. Prosodic exaggeration is thought to direct infants’ attention, modulate arousal and affect initially, and later fulfill more specific linguistic purposes such as lexical segmentation (Cooper & Aslin, 1994; Fernald, 1992). Formant exaggeration is phonetically associated with hyperarticulation (Johnson, Flemming & Wright, 1993), which may facilitate language learning by making the critical acoustic distinctions more salient and the phonetic categories more discriminable (Kuhl *et al.*, 1997). Supporting evidence indicates that vowel space in maternal speech is positively correlated with infants’ speech perception – mothers who tended to ‘stretch out’ their vowels had better-performing babies in phonetic discrimination (Liu, Kuhl & Tsao, 2003). Furthermore, computer models have demonstrated robust unsupervised learning of speech sounds based on IDS input (de Boer & Kuhl, 2003; Kirchhoff & Schimmel, 2005; Vallabha, McClelland, Pons, Werker & Amano, 2007). However, little is known about the neurobiological mechanisms that promote learning by exploiting the physical properties of IDS.

Brain research studies offer new insights into speech processing and language acquisition (Dehaene-Lambertz & Gliga, 2004; Kuhl *et al.*, 2008). Several

neurophysiological indices have been shown to be associated with IDS compared to adult-directed speech (ADS), including increased frontal cerebral blood flow (Saito, Aoyama, Kondo, Fukumoto, Konishi, Nakamura, Kobayashi & Toshima, 2007), increased frontal electroencephalography (EEG) power (Santesso, Schmidt & Trainor, 2007), and enhanced event-related potentials (ERPs) in the frontal-temporal-parietal recording sites (Zangl & Mills, 2007). The IDS-induced enhancement in neural activity may work jointly with arousal, attention and affect to strengthen auditory memory for phonological, syntactic and semantic categories. However, none of the previous infant studies controlled the acoustic parameters to determine the linguistic effects of formant exaggeration specific to IDS independent of the prosodic/affective effects primarily drawn from fundamental frequency (f_0) modifications that are also found in pet-directed speech (Burnham, Kitamura & Vollmer-Conna, 2002).

The present study utilized synthesized stimuli and high-density EEG to investigate neural coding of vowel formant exaggeration in infants. EEG records electrical potential signals from electrodes placed on the scalp. ERPs, which are derived from averaging EEG epochs time-locked to stimulus presentation, provide a direct noninvasive measure of postsynaptic activities with millisecond resolution, suitable for studying the online cortical dynamics of acoustic and linguistic processing (Dehaene-Lambertz & Gliga, 2004; Näätänen & Winkler, 1999). High-density EEG, which records data from 64 or more electrodes, additionally allows for reliable source estimation of high-quality ERP data (Izard, Dehaene-Lambertz & Dehaene, 2008; Johnson, de Haan, Oliver, Smith, Hatzakis, Tucker & Csibra, 2001; Reynolds & Richards, 2005). Furthermore, advancement in EEG time-frequency analysis has opened a new venue for studying the event-related oscillations (EROs) in infants (Csibra, Davis, Spratling & Johnson, 2000; Csibra & Johnson, 2007). EROs reflect time-varying neuronal excitability and discharge synchronization at different rhythms subserving communications between neuronal populations for various attentional, memory and integrative functions (Klimesch, Freunberger, Sauseng & Gruber, 2008). Studies have shown that delta (1–4 Hz) and theta (4–8 Hz) activities, among other EROs, are closely associated with linguistic processing (Radicevic, Vujovic, Jelicic & Sovilj, 2008; Scheeringa, Petersson, Oostenveld, Norris, Hagoort & Bastiaansen, 2009). It remains to be tested how formant-exaggerated speech affects neural activation and synchronization in the infant brain.

The experimental design followed a basic assumption in auditory neuroscience – the average response for repeated presentations of the same stimulus (or multiple instances of the same stimulus category) is equivalent to the neural representation of the stimulus (or the category), which codes its acoustic/perceptual features. An alternating block design with an equal

stimulus ratio was adopted for this purpose.¹ The design took into account developmental changes in infant ERPs. A number of studies have shown that speech perception and auditory ERPs change dramatically in the first year of life, and adult-like language-specific perception occurs by 6 months of age (e.g. Cheour, Ceponiene, Lehtokoski, Luuk, Allik, Alho & Näätänen, 1998; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Polka & Werker, 1994). The developmental changes in the latency, amplitude, polarity and scalp distribution of ERP responses have led to a better understanding of brain mechanisms that support phonetic processing and language learning. There are two salient auditory ERP components at this age, P150, a positive peak at approximately 150 ms, and N250, a negative peak at approximately 250 ms (Dehaene-Lambertz & Dehaene, 1994; Fellman, Kushnerenko, Mikkola, Ceponiene, Leipälä & Näätänen, 2004; Kushnerenko, Ceponiene, Balan, Fellman, Huotilaine & Näätänen, 2002; Novak, Kurtzberg, Kreuzer & Vaughan, 1989; Rivera-Gaxiola, Silva-Pereyra, Klarman, Garcia-Sierra, Lara-Ayala, Cadena-Salazar & Kuhl, 2007; Zangl & Mills, 2007).

Although an exact neurocognitive model is not available to test how exaggerated speech affects neural processing at the segmental level in infants, developmental studies have provided important details about the neural basis of speech perception early in life (Dehaene-Lambertz & Gliga, 2004; Kuhl *et al.*, 2008). Magnetoencephalography (MEG) data show that phonetic discrimination in infants at 6–12 months of age activates the inferior frontal and superior temporal regions in the left brain (Imada, Zhang, Cheour, Taulu, Ahonen & Kuhl, 2006). Functional magnetic resonance imaging (fMRI) data further reveal that activation for speech stimuli in the Broca's area can be found even in 3-month-old infants (Dehaene-Lambertz, Hertz-Pannier, Dubois, Mériaux, Roche, Sigman & Dehaene, 2006). The co-activation in Broca's and Wernecke's areas is thought to indicate perceptual-motor binding to promote speech learning. Consistent with imaging results, ERP studies suggest that both left and right auditory regions are

¹ ERP research in speech perception has primarily used the mismatch negativity (MMN) paradigm. The generation of MMN involves repeated presentation of the same stimulus (or stimulus category), which is occasionally replaced by another stimulus (or stimulus category) (Näätänen, Paavilainen, Rinne & Alho, 2007). The MMN is a powerful measure that reflects experience-dependent neural sensitivity in detecting a sound change. A different paradigm, which requires less recording time, presents stimuli in blocks with an equal occurrence ratio to study neural coding of speech stimuli independent of discriminatory sensitivity (e.g. Molfese & Molfese, 1980; Sharma, Marsh & Dorman, 2000; Tremblay, Friesen, Martin & Wright, 2003; Zangl & Mills, 2007). The alternating block design combined important features of both paradigms while keeping the stimulus presentation time relatively short for the infant participants (Supplemental Figure 1). Unlike the MMN experiment, the alternating block design shifted the focus to neural coding of the standard stimuli alone.

sensitive to coding acoustic/phonetic features of speech stimuli with striking similarities between infants and adults (Dehaene-Lambertz & Gliga, 2004). There exists limited evidence for left-hemisphere dominance for speech in infants, which may be attributable to a functional asymmetry of the auditory system in processing rapid acoustic transitions versus slow spectral changes (Poeppe, 2003; Zatorre & Belin, 2001). For instance, EEG and near-infrared spectroscopy data from newborn infants show bilateral activation for speech-like acoustic modulations and right-hemisphere dominance for acoustic modulations at a much slower rate (Telkemeyer, Rossi, Koch, Nierhaus, Steinbrink, Poeppe, Obrig & Wartenburger, 2009). The vowel stimuli in the present study did not contain rapid acoustic transition and thus provided an opportunity to test functional asymmetry for spectral processing in infants at 6–12 months of age.

The general hypothesis was that formant exaggeration would induce enhanced neural responses for speech processing in the infant brain. Specifically, the present study examined the effects of formant exaggeration in two ERP components (P150 and N250), two ERO bands (delta and theta), and three broad regions of interest (frontal, temporal/central, parietal) in both hemispheres. There were four closely related questions. First, at what time points, or in what ERP components, did the effect occur? Second, was the hypothesized effect mediated by differences in neural synchronization? Third, what cortical regions were affected? Fourth, did the data support early functional asymmetry for spectral processing of formant exaggeration? Answers to these questions would provide an initial account of the neural mechanisms responsible for the facilitative role of formant exaggeration in speech learning and acquisition.

Methods

Participants

Eighteen normally developing infants were recruited via advertisement. Informed parental consent was obtained in accordance with the procedures approved by the institutional Human Research Protection Program. The parents were paid \$20. The infants were full term with normal pregnancies and deliveries and no known auditory or visual problems. Both parents were native English speakers, and no immediate family members had a history of speech, language, or hearing deficits. Two infants did not complete the entire recording session. Four others were excluded due to excessive noise in their EEG data. The remaining 12 infants (seven girls, mean age = 9.2 months, range = 6.5–11.4 months) were included in this report. Infants at this age range were considered appropriate because their auditory N250 responses were known to be robust and sensitive to phonetic learning.

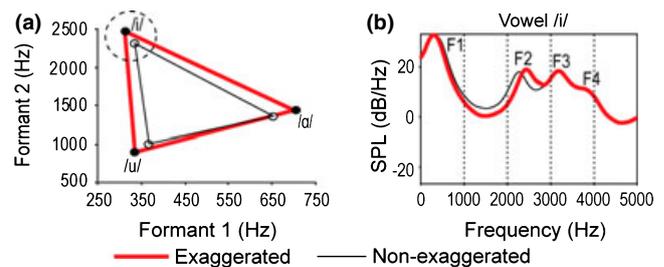


Figure 1 (a) Simulated vowel triangle expansion in the F1-F2 space for the three point vowels, /i/, /a/, and /u/. Only the /i/ sounds were used in this study. (b) Spectral overlay plot for the synthesized exaggerated and non-exaggerated /i/ stimuli.

Stimuli

The vowels were created with the Hlsyn program (Sensimetrics Corporation, USA) (Figure 1). Hlsyn allowed quality synthesis with direct control of formants and quasi-articulatory parameters based on the Klatt formant synthesizer (Hanson & Stevens, 2002). The sounds were 200 ms in duration (Supplementary materials, Sounds 1 and 2). The F1 and F2 parameters were chosen to simulate vowel space expansion based on previous studies (Hillenbrand, Getty, Clark & Wheeler, 1995; Kuhl *et al.*, 1997). Only the male /i/ sounds were used for this study.² Specifically, the center frequencies of F1 and F2 were 342 Hz and 2322 Hz for the non-exaggerated /i/. The exaggerated /i/ had F1 at 310 Hz and F2 at 2480 Hz. The f_0 , F3 and F4 were held constant at 138, 3000, and 3657 Hz for both sounds. Both vowels included a rise/fall time of 10 ms. They were resampled at 44.1 kHz and normalized with equal average RMS (root mean square) intensity.

Stimulus presentation

Infants sat on their parent's lap in an acoustically and electrically treated chamber (ETS-Lindgren Acoustic Systems). Stimulus presentation used the Evoke software (ANT Inc., The Netherlands). The sounds were played via a pair of loudspeakers (M-audio BX8a) placed at approximately 1.2 m in front and 40 degrees to the sides of the infant. The sound level was calibrated to be 65 dB SPL at the approximate location corresponding to the center position of the subject's head. Alternating blocks were presented for a total of 360 trials in 18 blocks, with each block consisting of 20 identical sounds. Block order was counterbalanced among the subjects.

² The male voice was chosen in consideration of naturalness of the synthesized vowel stimuli as judged by five adult native speakers of English. The Hlsyn software program is based on the Klatt formant synthesizer (Klatt & Klatt, 1990), and male-based speech synthesis has been used in previous infant speech perception studies (Kuhl, Tsao & Liu, 2003; Liu *et al.*, 2003). As pointed out by an anonymous reviewer, female speech that typically contains greater acoustic exaggeration could potentially produce a larger effect in infants than is reported in the present study.

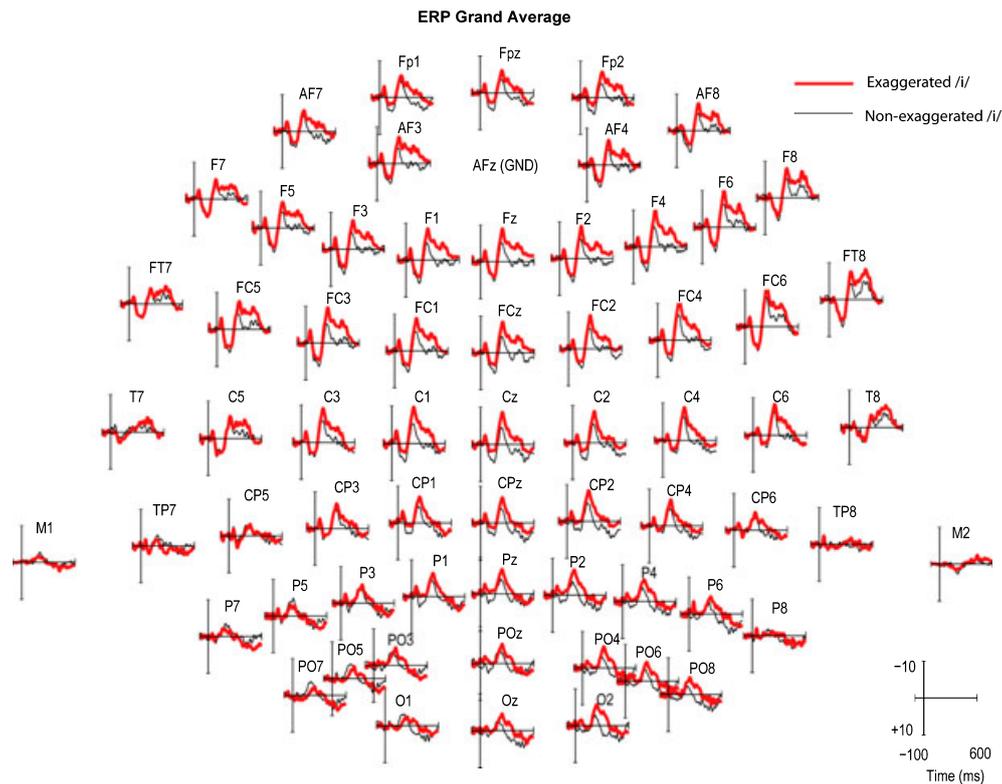


Figure 2 Grand average ERP plot (linked-mastoid reference) for the two speech stimuli in all 64 electrodes.

The offset-to-onset interstimulus intervals were randomized between 1.1 and 1.2 s, and the inter-block silence period was 5 s. The short blocks (each less than 30 s) and the relatively long randomized ISIs and inter-block silence served to reduce habituation of the important ERP components (Dehaene-Lambertz & Dehaene, 1994; Woods & Elmasian, 1986).

EEG recording

Continuous EEG was recorded (bandwidth = 0.016–200 Hz; sampling rate = 512 Hz) using the ASA-Lab system with REFA-72 amplifier (TMS International BV) and WaveGuard cap (ANT Inc., The Netherlands). Head circumference was measured for each infant to determine head size. The EEG cap used shielded wires for 65 sintered Ag/AgCl electrodes in the international 10–20 montage system and the intermediate locations (Figures 2, 3a, and 3b). The ground was positioned at AFz, and the default reference for the REFA-72 amplifier was the common average of all connected unipolar electrode inputs. In the EEG cap, each electrode was surrounded by a silicone ring to hold the conductive gel, which allowed the electrodes to be prefilled with gel and facilitated a smoother cap-fitting procedure. Adjustments on individual electrodes were made to keep impedances at or below 5 k Ω .

During the experiment, one or two research assistants sat in front of the infant, silently playing with toys to keep the infant's attention. With parental permission,

a muted cartoon movie was also played on a 20-inch LCD TV at 2.5 m away. The researcher in the control room communicated with the assistant and parent via intercom to coordinate the recording and initiate the session when the infant sat relatively still. A surveillance camera (Canon VB-C50iR) monitored the infant's behavior. The ASA-Lab automatically saved the online video in synchrony with EEG recording for later assessment of data quality and the infant's alertness level. When necessary, sound presentation and EEG recording would be paused until the infant sat relatively still again. The entire EEG session, including preparation, lasted approximately 40 minutes.

ERP waveform analysis

ERP averaging was performed offline in BESA (Version 5.2, MEGIS Software GmbH, Germany) following recommended guidelines (DeBoer, Scott & Nelson, 2007; Picton, Bentin, Berg, Donchin, Hillyard, Johnson, Miller, Ritter, Ruchkin, Rugg & Taylor, 2000). The EEG data were bandpassed at 0.5–40 Hz. The ERP epoch length was 700 ms, including a pre-stimulus baseline of 100 ms. The automatic artifact scanning tool in BESA was applied in two steps to detect bad electrodes and noisy signals. First, adjustments of rejection threshold parameters were made on every subject to inspect the effects on the entire recording session and the number of accepted trials. Second, the automatic rejection criteria were determined for all subjects. Epochs with a signal

level exceeding $150 \mu\text{V}$ from the segment baseline or a slew rate exceeding $75 \mu\text{V}/\text{ms}$ were rejected. A subject was excluded if the number of accepted trials for each stimulus condition did not meet the minimum of 40. As the focus of the study was on neural coding of the standard stimuli, the first stimulus in each block was excluded to avoid possible MMN elicitation from the alternating blocks. The average number of accepted trials was 58 for the exaggerated /i/ and 55 for the non-exaggerated /i/. Weighted averaging was calculated for the grand average to minimize influences of individual subjects with fewer trials. A caution against weighted averaging is that it might result in an undesirable bias towards drowsy infants, who could have a large number of uninformative epochs due to the decrease of ERP amplitude. For the present study, online observation and

video recordings indicated that there were no alertness problems for the infants (12 out of 18) included in grand averaging.

To keep consistency with the majority of previous infant ERP studies using speech stimuli (e.g. Ceponiene, Haapanen, Ranta, Näätänen & Hukki, 2008; Fellman *et al.*, 2004; Rivera-Gaxiola *et al.*, 2007; Zangl & Mills, 2007), the ERP data were re-referenced to linked mastoids. Further analyses were performed in Matlab (Version 7.6) with the EEGLAB software (Version 7.2.11) (Delorme & Makeig, 2004). To improve signal to noise ratio, nine electrode regions were defined for grouped averaging for each subject (Figure 3b): left frontal (LF, including F7, F5, F3, FT7, FC5, and FC3), middle frontal (MF, including F1, Fz, F2, FC1, FCz, and FC2), right frontal (RF, including F8, F6, F4, FT8,

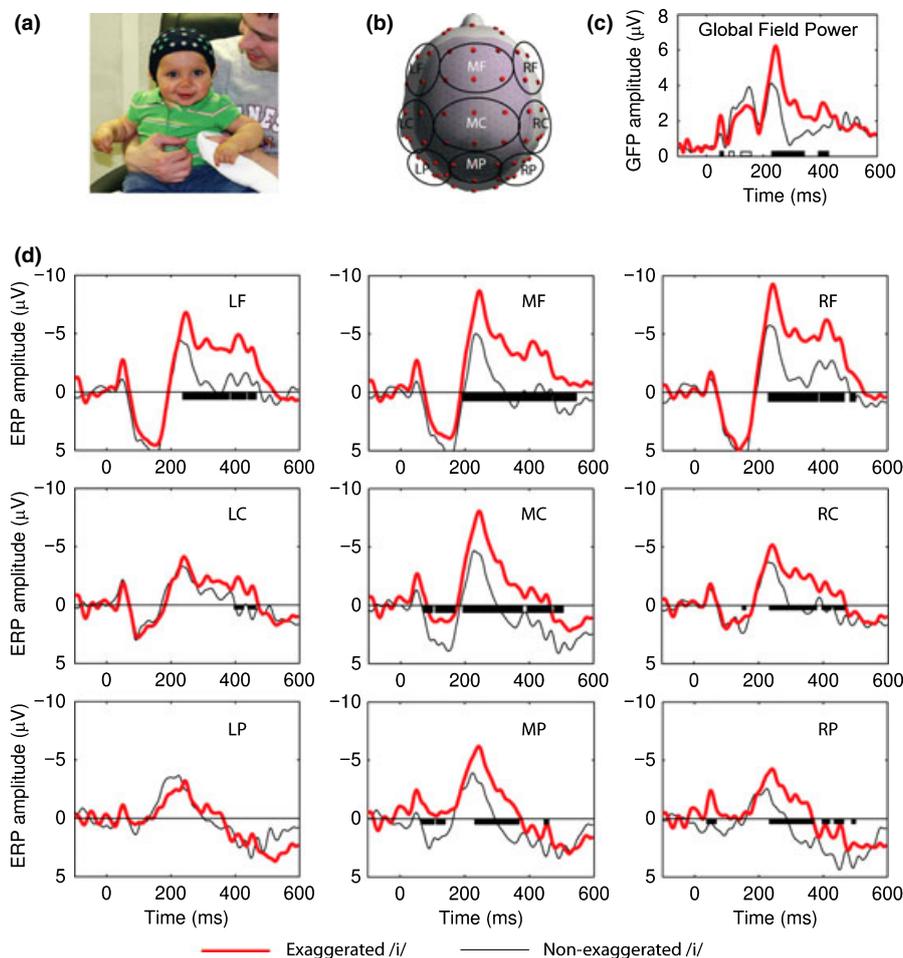


Figure 3 (a) Photo illustration of an 11-month-old infant wearing the 64-channel EEG cap during a break in the recording session. (b) Realistic head model with 64 standard electrode positions on. Nine grouped electrode sites were marked, representing frontal (F), central (C), and parietal (P) regions in the left (L), middle (M), and right (R) divisions. (c) Grand average global field power data for the exaggerated and non-exaggerated vowel stimuli. Significance for time-point-by-time-point comparison was shown by the black and white bars on the x-axis. The black bars on the x-axis showed time windows of significant enhancement for the exaggerated sound relative to the non-exaggerated sound, and the white bars showed time windows of significant reduction [$p < .01$]. (d) Grand average ERP data (linked-mastoid reference) at the nine regionally grouped electrode sites for point-by-point comparison between the two stimuli. The black bars on the x-axis showed time windows of significant negativity enhancement for the formant-exaggerated speech [$p < .01$]. A cautionary note is needed for proper interpretation as the electrode site effects are not equivalent to those of cortical regions in the brain.

FC6, and FC4), left central (LC, including T7, C5, C3, TP7, CP5, and CP3), middle central (MC, including C1, Cz, C2, CP1, CPz, and CP2), right central (RC, including T8, C6, C4, TP8, CP6, and CP4), left posterior (LP, including P7, P5, P3, PO7, PO5, and PO3), middle posterior (MP, including P1, Pz, P2, and POz), and right posterior (RP, including P8, P6, P4, PO8, PO6, and PO4) (Schneider, Debener, Oostenveld & Engel, 2008). To derive an unbiased estimate independent of electrode selection, global field power (GFP) was calculated for comparison by computing the standard deviation of the amplitude data across the 64 electrodes at each sampling point (Hamburger & van der Burgt, 1991; Lehmann & Skrandies, 1980). Repeated-measures ANOVA tests were performed on the mean amplitudes of a 20 ms interval around peaks of interest. The peak search windows (60–200 ms for P150 and 200–500 ms for N250) were confirmed by inspection of the grand mean ERP overlay plots and topography. Peak-to-peak (P150–N250) values were also calculated for statistical comparison. The within-subject factors were stimulus type (exaggerated and non-exaggerated), hemisphere (left and right) and electrode region (frontal, central, and parietal). Where appropriate, either Bonferroni or Greenhouse-Geisser correction was applied to the reported *p*-values.

There have been controversies regarding the use of linked mastoids in ERP research (Picton *et al.*, 2000). The topographical voltage maps of linked-mastoid reference may be incorrect because the mastoids are not neutral electrodes for auditory stimuli (Dehaene-Lambertz & Gliga, 2004; Michel, Murray, Lantz, Gonzalez, Spinelli & Grave de Peralta, 2004). To examine the effects of reference selection, ERP waveform analysis using common average reference was also performed. The common average reference method assumes that the average of all recording electrodes on a volume conductor is approximately neutral. However, this assumption is valid only with accurate spatial sampling of the scalp field, which requires a sufficient number of electrodes with full coverage of the head surface. The average of 64 electrodes with the standard 10-10 montage might be insufficient to qualify as a truly neutral reference (Dien, 1998; Yao, Wang, Oostenveld, Nielsen, Arendt-Nielsen & Chen, 2005). To help evaluate the differences between the two reference methods, the ERP data were further transformed into reference-free current source density (CSD) estimates (Kayser & Tenke, 2006; Perrin, Pernier, Bertrand & Echallier, 1989). In this approach, spherical spline surface Laplacian transform was applied to identify locations and relative magnitudes of current sources and sinks. The CSD waveforms were submitted to unrestricted Varimax-rotated temporal principal components analysis (PCA) based on estimating the source-current covariance matrix from the measured-data covariance matrix (Kayser & Tenke, 2006). The CSD estimates served to sharpen ERP scalp topography by

eliminating the volume-conducted contributions and the dependence on reference.

Time-frequency analysis

Time-frequency representations (TFRs) were derived for evoked EROs in the frontal (MF), central (MC) and parietal (MP) electrode regions using continuous wavelet transform (CWT) in Matlab (Csibra & Johnson, 2007; Samar, Bopardikar, Rao & Swartz, 1999). CWT was applied to the averaged ERP signals (referenced to linked mastoids) with complex Morlet wavelets (bandwidth = 1 Hz, center frequency = 0.5 Hz) in Matlab. Morlet scalograms were plotted using the absolute values of the squared coefficient of CWT, and the frequency values on the plots were converted from the normalized scale vector. The power data for delta (1–4 Hz) and theta (4–8 Hz) bands at the MF, MC, and MP electrode sites were subject to further analysis to examine the temporal evolution of formant exaggeration.

Temporal evolution analysis

Two-tailed time-point-to-time-point *t*-tests were performed to obtain accurate information about the temporal evolution of significant differences between the exaggerated and non-exaggerated stimuli (Guthrie & Buchwald, 1991). Guthrie and Buchwald's method requires three pieces of information to assess ERPs' temporal evolution at a chosen significance level: (1) sample size (*N*, representing the number of subjects), (2) number of sampling time points (*T*), and (3) the autocorrelation value (ϕ). This parametric test method relies on the computation of minimum number of consecutive sample points that need to show significant differences, which depends on the autocorrelation estimates and the total number of sample points in the ERP data. The analysis was applied to the GFPs, ERPs, CSDs and TFRs. For the present study, the data for an entire epoch were decimated and assessed at two segments (0–200 ms covering P150 and earlier components, and 200–600 ms covering N250 and later components) in order to be consistent with Guthrie and Buchwald's specific recommendations for the three parameters. For *N* = 12, the ϕ values obtained from all IDS-versus-ADS comparisons in the study were in the range of 0.38 to 0.74, corresponding to at least 5–9 consecutive sample points at the significance level of 0.01. For a conservative estimate, an interval would not be considered to differ significantly in this study unless at least nine consecutive sample points (approximately 18 ms) showed significant differences at the level of 0.01.

Source localization analysis

Minimum norm estimation (MNE L2-norm) was applied to the averaged ERP data (Hämäläinen & Ilmoniemi, 1994; Izard *et al.*, 2008). The MNE analysis approximated

the current source space using hundreds of prefixed, discrete and distributed dipoles directly within the cortex and searched for the optimal estimate with the smallest norm to explain the measured ERP signals. Unlike CSDs that are second-order spatial derivatives representing the current sources and sinks on the scalp, MNEs are modeled as true reference-free estimates of cortical current activities. The implementation for the present study included the following steps:

1. The infant electrode montage was calculated by scaling the standard positions for the WaveGuard EEG cap to fit the average head circumference of the 12 infant subjects.
2. A three-shell layer model was used to approximate the infant head (Reynolds & Richards, 2005). To improve the focality and reliability of the source activities, the MNE procedure included both depth weighting and spatio-temporal weighting (Dale & Sereno, 1993; Lin, Witzel, Ahlfors, Stufflebeam, Belliveau & Hämäläinen, 2006) to avoid bias towards superficial sources. Noise regularization used the lowest 15% values, and baseline noises were weighted by the average over the 64 electrodes. The entries in the main diagonal of the noise covariance matrix were equally proportional to the average noise power over all channels.
3. The total activity at each source location was computed as the root mean square of the source activities of its three components. The total activity solutions were projected to the standard realistic brain model in BESA. The current source data for 750 prefixed locations (x, y, z coordinates covering the entire brain space) at all latencies (358 sample points for the -100 – 600 ms epoch) were further analyzed for temporal and spatial interpretations.
4. In the temporal analysis, the total MNE activities in each hemisphere were summed at each time point for each stimulus. The MNE differences between the two stimuli at each sample point were then subjected to two-tailed z -test relative to the baseline mean and variance. To examine regional contributions to the total MNE activities, standard anatomical boundaries in the Talairach space were used to define the spatial masks for each region of interest (ROI) in the brain space (Lancaster, Woldorff, Parsons, Liotti, Freitas, Rainey, Kochunov, Nickerson, Mikiten & Fox, 2000). The anatomical ROIs allowed a crude estimation for frontal, temporal, and parietal activities in the two hemispheres separately (Zhang, Kuhl, Imada, Kotani & Tohkura, 2005).
5. In the spatial analysis, temporal integration was performed. The MNE differences between the two stimuli at each location and each time point were converted to z -scores relative to the distribution of baseline activities. The average z -scores were then calculated for the two selected time windows (0 – 200 ms and 200 – 600 ms) at each source location and plotted using Matlab visualization functions.

Results

ERP waveform results for linked-mastoid reference

The ERP data for the vowel stimuli showed clear auditory N50, P150 and N250 responses and traceable deflection for N450 in the sustaining negativity in the frontal-central sites (Figures 2, 3). The P150-N250 complex was identified in all 12 infant subjects. A clear N50 response was observed in 10 out of the 12 subjects. Compared with the non-exaggerated /i/, the exaggerated /i/ elicited more negative ERP responses in both the early time window (prior to 200 ms) dominated by P150 and the late window (subsequent to 200 ms) dominated by N250. Point-to-point t -test on the global field power (GFP) showed significantly larger ERPs for the exaggerated sound in three components, N50 (44–62 ms), N250 (228–356 ms), and sustaining negative activities (392–431 ms) [$p < .01$] (Figure 3c). Significantly smaller GFP values were found at 78–97 ms and 136–166 ms for the exaggerated sound, indicating relatively more negative responses in the positive window.

Repeated-measures ANOVA results were obtained from P150, N250, and peak-to-peak (P150-N250) amplitude data on the LF, LC, LP, RF, RC, and RF sites. For P150, there were significant main effects of stimulus type [$F(1, 11) = 5.28, p < .05$] and region [$F(2, 22) = 11.30, p < .01$]. For N250, significant effects were observed for stimulus type [$F(1, 11) = 20.45, p < .001$], hemisphere [$F(1, 11) = 7.66, p < .01$] and region [$F(2, 22) = 5.75, p < .01$], and hemisphere-by-region interaction [$F(2, 22) = 4.99, p < .05$]. The P150 and N250 responses got progressively closer to baseline in the frontal-central-parietal direction, and the right hemisphere showed dominance in N250. In the peak-to-peak (P150-N250) analysis, significant effects were found for stimulus type [$F(1, 11) = 29.97, p < .001$] and region [$F(2, 22) = 24.38, p < .001$], suggesting that the effects in P150 and N250 were not due to an overall shift of the P150-N250 complex.

Regional breakdown of the point-to-point comparisons between exaggerated and non-exaggerated /i/s showed detailed contributions from the frontal, central and parietal recording sites at different time intervals (Figures 3d and 6a). Significant effects were observed for N50 (at RP), P150 (at MC, MP, and RC), and N250 (at LF, MF, MC, MP, RF, RC, and RP) as well as for N450 (at LF, LC, MF, MC, MP, RF, RC and RP).

ERP waveform results for common average reference

The ERP data with common average reference showed clear auditory N50, P150, N250 and N450 responses in the frontal electrode sites (LF, MF and RF) (Figures 4 and 6b). Unlike the ERPs with linked-mastoid reference, the parietal sites (LP, MP, RP) uniformly showed polarity reversal to the frontal sites and the vertex,

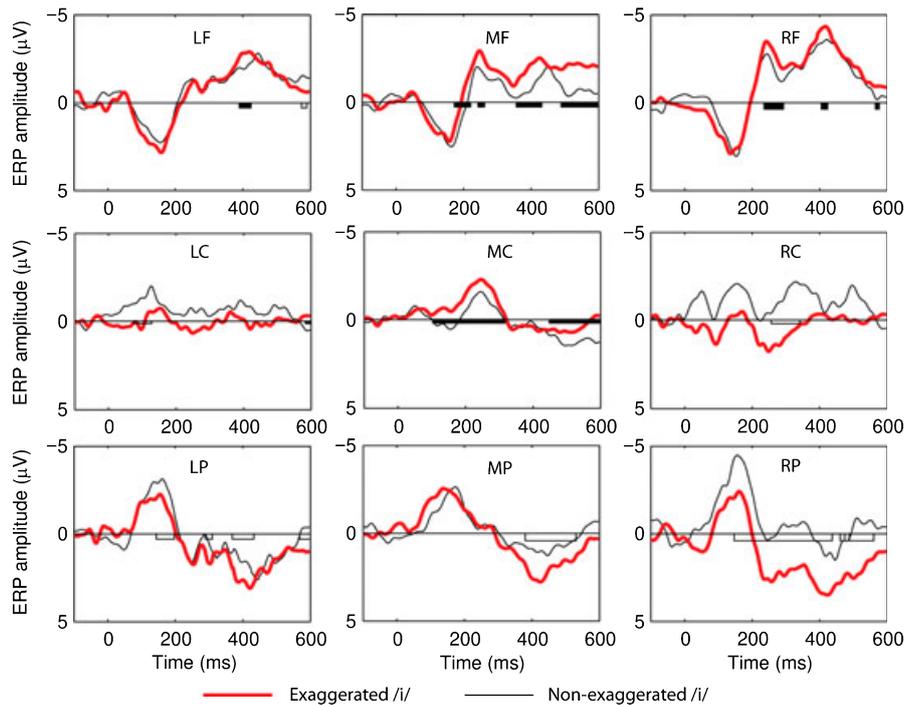


Figure 4 Grand average ERP data (common average reference) at the nine regionally grouped electrode sites for point-by-point comparison between the two stimuli.

suggesting the existence of dipole current sources in the left and right temporal cortices. The scalp potential distribution pattern created problematic averaging for the LC and RC sites because the co-occurring sources and sinks at the same sampling time point would cancel each other out in these two electrode groups.

Repeated-measures ANOVA showed similarities and differences between the two reference methods. For P150 amplitudes, there were significant effects for stimulus type [$F(1, 11) = 9.86, p < .01$] and region [$F(2, 22) = 13.49, p < .001$]. For N250 amplitudes, there were main effects of region [$F(1, 11) = 32.98, p < .0001$], hemisphere [$F(1, 11) = 5.28, p < .05$], and stimulus-by-region interaction [$F(1, 11) = 17.25, p < .001$]. As in the linked-mastoid reference analysis, the right hemisphere was dominant for N250.

Point-to-point comparisons between exaggerated /i/ and non-exaggerated /i/ showed enhanced N250 in MF, RF, and MC. However, there was no such N250 effect in LF. Contrary to the linked-mastoid reference, the significant differences were more extensive in the parietal electrodes than the frontal electrodes. The N250 enhancement showed up as increased positivity in LP and RP sites due to polarity reversal. Significantly more negative responses prior to N250 were observed in the middle line frontal and central sites (MF and MC), and this effect was observed with reversed polarity in left temporal-parietal sites (LC and LP). Significant enhancement in sustaining negativity after N250 was found in LF, MF, MC and RF with polarity reversal in LP, MP and RP sites. No significant differences were found in any electrode sites for the N50 response.

CSD waveform results

The reference-free CSD estimates showed polarity reversal in the frontal-parietal sites, which was similar to ERP data with common average reference. The existence of dipole current distribution in the bilateral temporal regions produced problematic averaging for LC and RC electrodes (Figures 5 and 6c). Unlike the ERP data, the CSD estimates closely approximated localization of cortical activities, and did not show strong P150 and N250 responses in the midline electrodes.

Repeated-measures ANOVA results showed consistent significant effects in P150 for stimulus type [$F(1, 11) = 11.27, p < .01$] and region [$F(2, 22) = 9.76, p < .01$]. There were main effects in N250 for region [$F(1, 11) = 23.37, p < .0001$], hemisphere [$F(1, 11) = 5.28, p < .05$], and stimulus-by-region interaction [$F(1, 11) = 27.51, p < .001$]. Point-to-point comparisons showed significant contributions to P150 differences in MC, LP, and MP regions, N250 differences in LF, RF, MC, RC, LP, and RP regions, and sustaining negativity differences in MF, RF, MC, LP and RP regions.

Topography and TFR results

The grand average ERP and CSD overlay plots and topographical maps for the two vowels illustrated the existence of N50, P150, N250 and N450 components (Figures 6a, 6b, 6c). Despite the differences in the statistical comparisons among linked-mastoid reference, common average reference and the reference-free CSDs, the P150 amplitudes were consistently maximal at

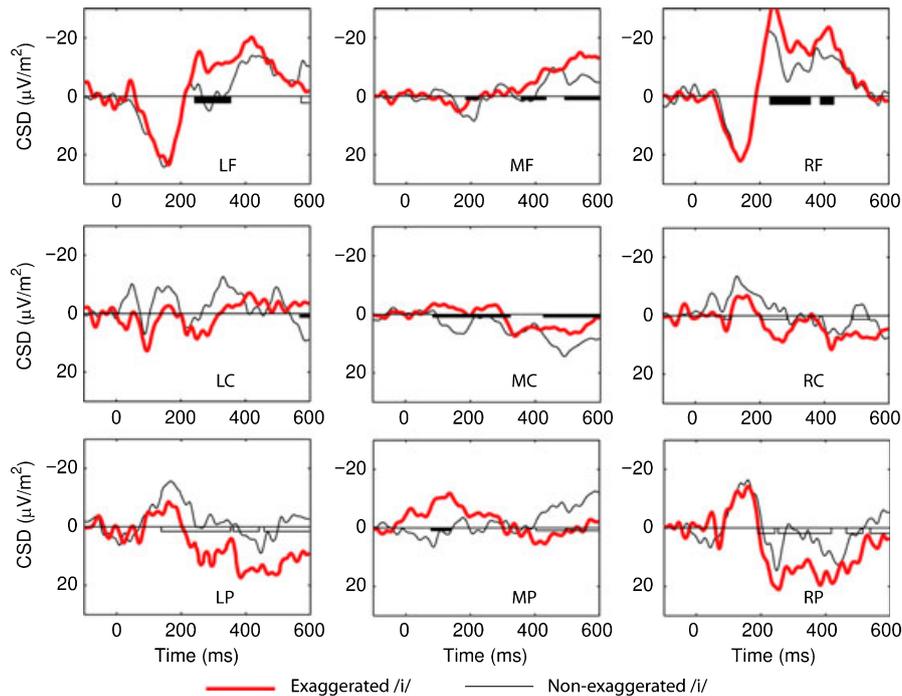


Figure 5 Grand average CSDs (current source density) at the nine regionally grouped electrode sites for point-by-point comparison between the two stimuli.

bilateral frontal electrode sites, and the N250 was dominant in temporal-parietal sites with functional asymmetry in favor of the right hemisphere. The N250 enhancement effect was clearly visible at temporal-parietal sites for exaggerated /i/ relative to the non-exaggerated /i/.

The dominant evoked EROs (linked-mastoid reference) were found in the delta (1–4 Hz) and theta (4–8 Hz) bands corresponding to the P150–N250 complex in time with linked-mastoid reference (Figure 6d). Differences between the two stimuli were observed with stronger ERO power in favor of the exaggerated sound. The overall ERO power got progressively weaker in the frontal-central-parietal direction. Strong delta and theta activities were present at the frontal and central electrode sites, and the parietal site predominantly showed delta activity. In point-to-point comparisons, significantly enhanced delta activity was found for the exaggerated /i/ sound at all three sites (154–425 ms at MF, 152–388 ms at MC, and 152–502 ms at MP) [$p < .01$]. Differences in theta band power were significant only in the frontal and central sites (162–281 ms at MF, and 201–283 ms at MC) [$p < .01$].

MNE results

The MNE plots showed activation in frontal, temporal, and parietal regions for the two vowel stimuli in both hemispheres (Figures 7a and 7b). Point-to-point comparisons showed significantly enhanced total cortical activity (N250 and sustaining activity following N250) for the exaggerated sound in both left (231–440 ms, 537–571 ms) and right hemispheres (190–482 ms,

500–600 ms) [$p < .01$]. As expected, temporal regions were the primary contributor to the enhancement in N250 and sustaining activities. There were also contributions from frontal and parietal regions in the two hemispheres. Unlike the waveform data, the total MNE amplitude data showed no significant difference in P150 for either hemisphere. The lack of significance in total activity for P150 could arise from regularization of cortical source diffusion and depth weighting in the MNE calculation. Significant regional P150 differences were found bilaterally in the frontal MNE activities [$p < .01$]. Although the peak amplitude for N250 appeared to show right-hemisphere dominance, the mean MNE differences (total MNE activity between the two stimuli in the window of 270–450 ms) were slightly larger in the left hemisphere than in the right.

The significant MNE differences for the early window (0–200 ms) were localized primarily in the inferior frontal area in both hemispheres. In the late time window (200–600 ms), the significant differences were found in the left inferior frontal cortex as well as right anterior temporal cortex extending posteriorly to superior temporal, inferior frontal and middle frontal regions (Figure 7c).

Discussion

Speech scientists have long stressed the importance of formant exaggeration in infant-directed speech for phonetic learning (Burnham *et al.*, 2002; Kuhl *et al.*, 1997). The ERP waveforms (including CSD waveforms), TFRs,

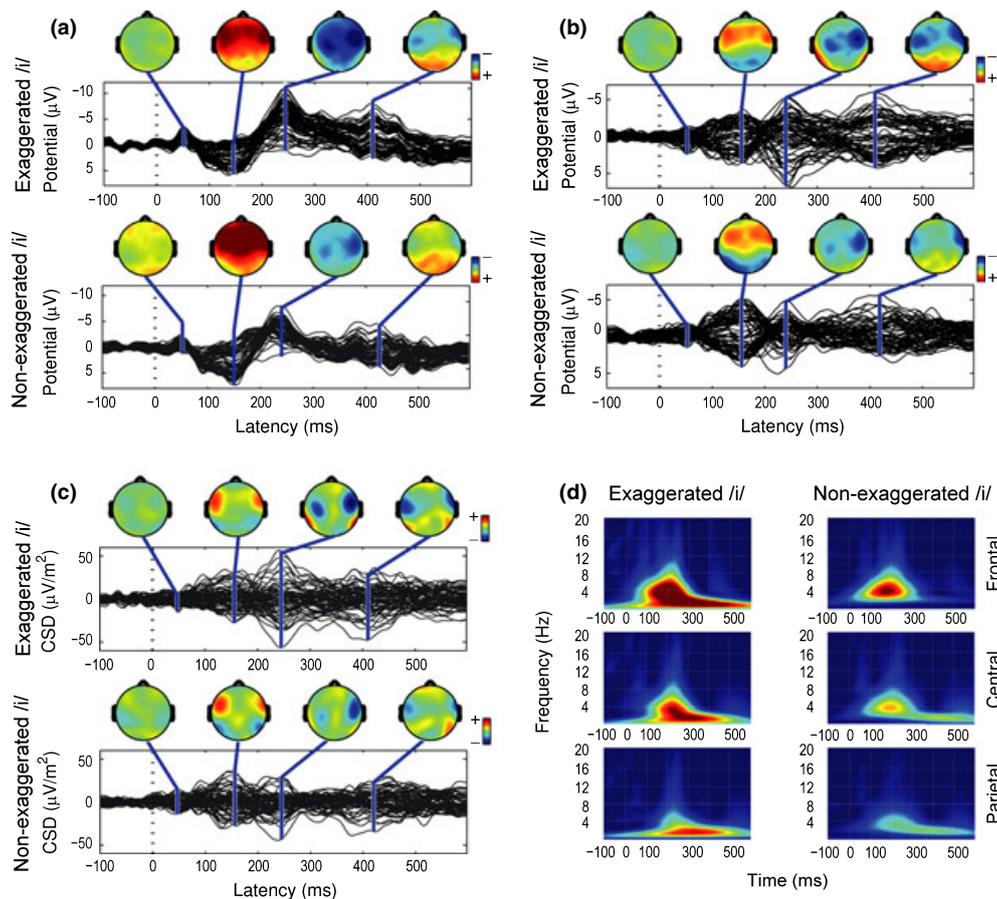


Figure 6 (a) Grand average ERP overlay plots of 64 electrodes (linked-mastoid reference) and topography maps of N50, P150, N250 and N450 responses for the two vowel stimuli. (b) Grand average ERP overlay plots (common average reference) and topography maps for the two stimuli. (c) Grand average CSD overlay plots and topography maps for the two stimuli. (d) Grand average Morlet scalograms at the MF (frontal), MC (central), and MP (parietal) sites for the exaggerated and non-exaggerated /i/ sounds in the study.

and MNE data here provided three lines of evidence in support of this view. Despite striking differences in ERP waveforms due to reference choice, significant enhancement in N250 and sustaining activity following N250 for exaggerated speech was confirmed in all the analyses. The reduced P150 effect, on the other hand, was not consistently found. Unlike N250, the early P150 component presumably reflected acoustic mapping of the spectral differences between the stimuli (Rivera-Gaxiola *et al.*, 2007). This functional distinction was partly supported by the scalp distribution of the components in all three topographical calculations using linked-mastoid reference, average reference, and the reference-free CSD approach. The P150 was dominant in the frontal sites, and the N250 extended posteriorly from frontal to temporal-parietal electrode sites. The timing and scalp distribution of the enhanced negativity in the 200–600 ms window were consistent with the notion that the N250 and sustaining negative responses are linked with phonetic and lexical processing in infants at 6 months of age or older (Mills, Prat, Zangl, Stager, Neville & Werker, 2004; Rivera-Gaxiola *et al.*, 2007; Zangl & Mills, 2007). An alternative interpretation is that the P150 and N250 responses do not necessarily serve the strict bifurcation

of auditory vs. linguistic processing. Rather, these two components co-occur and behave similarly in many experimental situations, and may thus reflect connected processes. In line with both of these interpretations, a missing or diminished N250 was found to be associated with lower level of cognitive and linguistic development and diverted central auditory processing (Ceponiene *et al.*, 2002; Fellman *et al.*, 2004; Tonnquist-Uhlen, 1996).

Differential patterns of neural activity for IDS and ADS have been reported in previous infant studies (Saito *et al.*, 2007; Santesso *et al.*, 2007; Zangl & Mills, 2007). Saito and colleagues employed near-infrared spectroscopy in examining neonates' responses to naturally spoken sentences in the two speech styles. They found that IDS increased frontal activation in neonates, which was mainly attributable to the prosodic exaggeration of IDS and its socio-affective impact. Santesso *et al.* showed that in 9-month-old infants, the overall frontal activation in terms of EEG power was linearly related to affective intensity of natural sentences spoken in IDS. Zangl and Mills compared ERPs for words spoken in IDS and ADS in 6- and 13-month-old infants and found larger N600–800 responses to IDS than to ADS in both

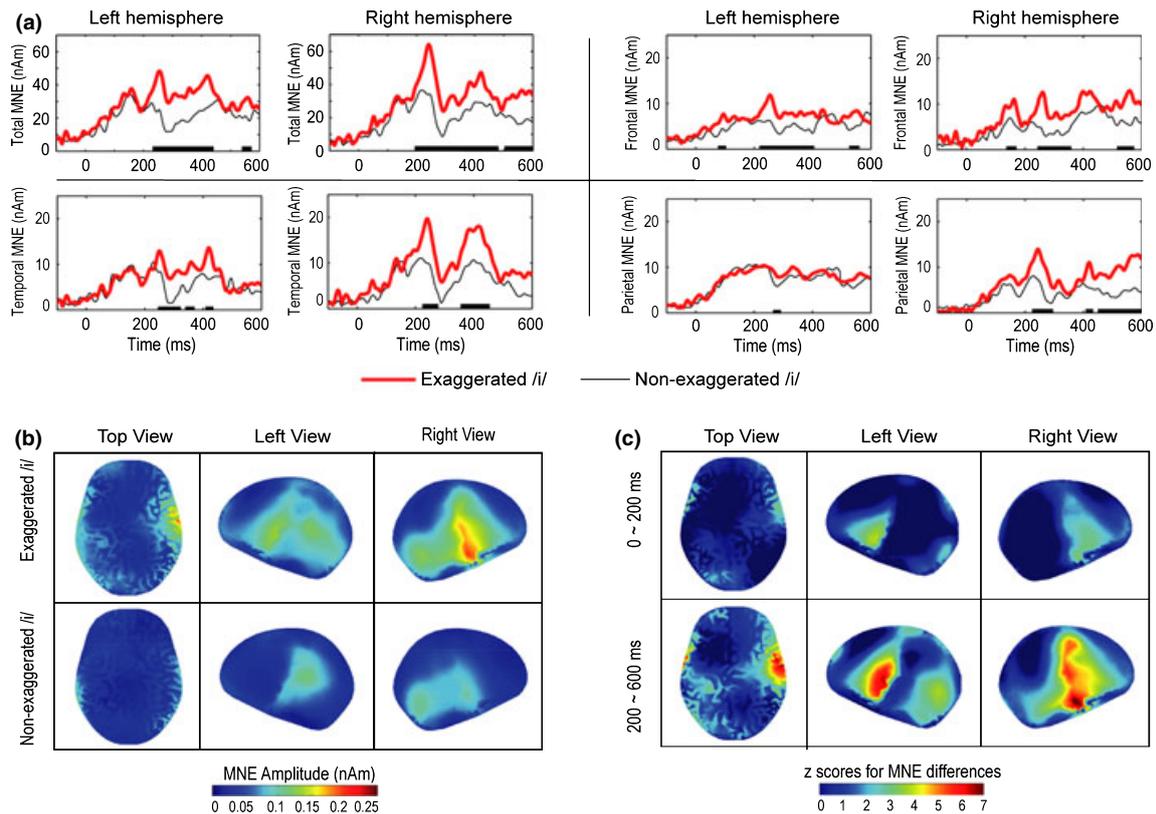


Figure 7 (a) Temporal evolution of total MNE activity and regional (frontal, temporal and parietal) MNE activities for the two stimuli in the two hemispheres. The black bars on the x-axis show time windows of significant MNE enhancement for the exaggerated speech relative to baseline current activities [$p < .01$]. (b) Spatial localization in top, left and right views of the MNE estimates (in nAm) for the two vowel stimuli. The MNE activities were integrated over 0–600 ms at each prefixed spatial location. (c) Spatial localization in top, left and right views of the MNE differences between the two vowel stimuli. The MNE differences were converted into z-scores relative to baseline and integrated over two time windows (0–200 ms and 200–600 ms).

age groups. In the older infants, familiar words additionally showed enlarged N200–400 response to IDS. Given that the IDS stimuli in the previous study were significantly longer in duration and higher in fundamental frequency, maximum pitch, and frequency range than ADS stimuli (Zangl & Mills, 2007), the increased brain activity for IDS would presumably reflect a composite effect of both prosodic and linguistic exaggerations. By controlling acoustic exaggeration other than formants in IDS and ADS, the new ERP data here demonstrated that formant exaggeration alone at the segmental level could produce significant enhancement in neural activation in 6–12-month-old infants, which may serve to strengthen associations between phonetic processing and word learning (Swingley, 2009).

The mechanism for the observed enhancement in N250 and sustaining negativity appears to rely on neural synchronization of evoked EROs time-locked and phase-locked to stimulus presentation. In the literature, the adult theta activity has been linked with arousal/orienting responses and working memory of verbal stimuli (Basar, Basar-Eroglu, Karakas & Schürmann, 1999; Hwang, Jacobs, Geller, Danker, Sekuler & Kahana, 2005; Klimesch, Hanslmayr, Sauseng, Gruber, Brozinsky, Kroll, Yonelinas & Doppelmayr, 2006; Scheeringa *et al.*,

2009; Summerfield & Mangels, 2005). In infants, delta (1–4 Hz) and theta (4–8 Hz) activities are both affected by linguistic processing with increased theta power for affective speech (Orekhova, Stroganova, Posikera & Elam, 2006; Radicevic *et al.*, 2008; Santesso *et al.*, 2007). As the pitch level was controlled in the present study, the observed increases in delta activity at frontal-central-parietal sites, as well as in theta activity at frontal-central sites, could not be due to prosodic processing. Rather, it could be a composite effect of attentional and phonetic encoding processes in response to the acoustically more salient and phonetically more distinct speech (Kuhl *et al.*, 1997). As attention was not controlled in the present study, it remains to be tested whether formant exaggeration alone makes speech more attractive to infant listeners.

The MNE differences between the stimuli revealed a bilateral cortical neural network sensitive to formant exaggeration, including the Broca's area in the left brain and frontal-temporal-parietal regions in the right. Broca's activation for speech processing has been reported in imaging studies of infants at 3–12 months of age (Dehaene-Lambertz *et al.*, 2006; Imada *et al.*, 2006), suggesting the existence of early perceptual-motor binding in support of language acquisition. The present

MNE data further indicate that formant-exaggerated speech leads to enhanced Broca's activation, which may drive speech learning via interactions with the perceptual-motor system involving temporal, frontal, and parietal cortices in both hemispheres. It is interesting to note that the infant MNE activation patterns for passive listening to speech show striking resemblance to adult fMRI data during passive listening to music (Lahav, Saltzman & Schlaug, 2007). Adult imaging research has also shown that auditory listening alone can recruit production-related regions including Broca's area (Love, Haist, Nicol & Swinney, 2006; Meyer, Steinhauer, Alter, Friederici & von Cramon, 2004; Skipper, Nusbaum & Small, 2005; Wilson, Saygin, Sereno & Iacoboni, 2004). The adult data were thought to reflect more general mnemonic and integrative functions for the Broca's area in making associations between motor actions for sound generation (not just speech) and the acoustic product. However, passive listening to nonsense syllables does not reliably elicit inferior frontal activation in adults (Zhang, Kuhl, Imada, Iverson, Pruitt, Stevens, Kawakatsu, Tohkura & Nemoto, 2009; Zhang *et al.*, 2005). As no motor component of speech is measured for comparison in the present design, it remains purely speculative that passive listening to speech might elicit motor activities in the developing minds to mediate phonological acquisition.

The ERP, CSD and MNE data all indicated greater involvement of the right hemisphere for the N250 effect than the left. This result was consistent with a recent study that showed early functional asymmetry of spectral processing in newborns (Telkemeyer *et al.*, 2009). There is a growing literature relating the right hemisphere with speech processing at the prelexical and paralinguistic levels (e.g. Bristow, Dehaene-Lambertz, Mattout, Soares, Gliga, Baillet & Mangin, 2009; Homae, Watanabe, Nakano, Asakawa & Taga, 2006; Scott & Wise, 2004; Simos, Molfese & Brenden, 1997). However, the laterality result directly contradicted previous findings about left-hemisphere dominance in significantly enhanced N200–400 and N600–800 responses for familiar words spoken in IDS relative to ADS (Zangl & Mills, 2007). The laterality inconsistency can be explained by the functional asymmetry model – spoken words involve fine-scale temporal processing of the rapid acoustic transitions in the left brain, and processing steady spectral cues in simple vowel stimuli primarily depends on the right brain (Poeppel, 2003; Zatorre & Belin, 2001). Nevertheless, this model did not specify the time course of functional asymmetry or the time course of interactions of cortical regions in auditory processing. A simple extrapolation would predict the same pattern of functional asymmetry regardless of the time course of brain activities, which was not supported by the current results. Further research is necessary to determine left/right functional asymmetries at different cortical regions and in different time windows and how asymmetry in brain activation varies as a function of stimulus properties, task variables, and subject characteristics.

The reference-dependent and reference-free approaches in the present study showed similarities as well as striking differences. The ERP research field has yet to adopt one standard solution regarding the choice of reference. Caution must be used in interpreting ERP results with different reference methods (Dien, 1998; Yao *et al.*, 2005). The topographical map for common average reference was similar to the CSD map in terms of the polarity reversal pattern. As all electrical activity produces dipolar fields, measurements from the two sides of the dipolar activity will always be negatively correlated. It is noteworthy that polarity reversal in the temporal-parietal electrodes relative to frontal electrodes could potentially cause problems in channel grouping and interpretation. Compared with common average reference, the linked-mastoid reference appears to produce biophysically unrealistic unipolar voltage fields. Although linked mastoid reference was quite popular in the past, it is recommended that researchers should switch to more progressive approaches by adopting the common average reference in future studies. While the CSD and MNE solutions have the advantages of being reference-free, these methods are highly susceptible to noise influence and thus technically challenging to implement when analyzing individual subjects' data, especially those of infants where there tends to be more noise.

As children learn to speak only the language(s) that they are exposed to, defining the role of input and the neurobiological mechanisms enabling this feat is central to our understanding of the perceptual and computational processes that adaptively shape both the developing brain and the language outcome. There is cumulative evidence that the acoustic and linguistic modifications in IDS have important functions in the acquisition of phonology and grammar (Burnham *et al.*, 2002; Liu *et al.*, 2003; Morgan & Demuth, 1996; Werker, Pons, Dietrich, Kajikawa, Fais & Amano, 2007). The present results add a neural-level account of how formant exaggeration in speech alters infants' brain activities for phonetic processing. This account is not without its limitations in explaining the role of formant exaggeration in language acquisition. Research has shown that not all aspects of acoustic exaggeration in IDS necessarily aid speech discriminability or learning (Trainor & Desjardins, 2002). As the distributional and statistical properties of language input are embedded within an interactive social learning environment (Meltzoff, Kuhl, Movellan & Sejnowski, 2009), it seems unlikely that any single property of IDS is indispensable to normal language development.

Of particular interest to theory and practice is that the effects of enriched language exposure, including formant exaggeration, are not limited to infancy. IDS-based input manipulation is conceptualized to be an agent of neural plasticity regardless of age or experience (Zhang *et al.*, 2009). The benefits of various input manipulations have been demonstrated in infants, children, and adults with or without learning disabilities (Bradlow, Kraus &

Hayes, 2003; Kuhl *et al.*, 2003; Tallal, 2004; Zhang *et al.*, 2009). Given that early brain measures have predictive power for later language skills (Kuhl *et al.*, 2008; Molfese, 1989), more developmental studies are needed to delineate the role of language input in the social context of language acquisition or effective intervention. In particular, an experimental design focusing on the different spectral and temporal aspects of IDS is necessary to build a better understanding of cortical speech processing and functional asymmetry. Both speech stimuli and nonspeech control can be applied to further investigate the effects of acoustic versus phonetic processing in populations of specific ages and neurological conditions (Dehaene-Lambertz & Gliga, 2004).

In summary, the present study examined the effects of formant exaggeration on cortical speech processing in infants at 6–12 months of age. Despite methodological differences, there was significant enhancement in N250 with right-hemisphere dominance in all reference-dependent and reference-free analysis approaches. Time-frequency analysis indicated increased neural synchronization for processing formant-exaggerated vowel stimuli in the delta band at frontal-central-parietal electrode sites as well as in the theta band at frontal-central sites. Minimum norm estimates further revealed a bilateral cortical neural network (frontal, temporal and parietal regions) in the infant brain sensitive to formant exaggeration, which may facilitate learning via cortical interactions in the perceptual-motor systems. Although there was limited support for the early functional asymmetry for spectral processing of formant exaggeration in the right hemisphere, hemispheric laterality may vary depending on the time course of neural activation.

Acknowledgements

This study was supported by funding sources to YZ, including two Brain Imaging Research Awards from the College of Liberal Arts (CLA) and the Grant-in-aid Program at the University of Minnesota. The first author would like to thank three anonymous reviewers for revision suggestions, Drs Matti Hämäläinen, Iku Nemoto, and Masaki Kawakatsu for technical discussions, and CLA Associate Deans, Jo-Ida C. Hansen and Jennifer Windsor, for support.

References

- Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *Journal of the Acoustical Society of America*, **119**, 1636–1647.
- Basar, E., Basar-Eroglu, C., Karakas, S., & Schürmann, M. (1999). Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG? *Neuroscience Letters*, **259**, 165–168.
- Bradlow, A.R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, **46**, 80–97.
- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J.F. (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, **21**, 905–921.
- Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new pussycat: on talking to animals and babies. *Science*, **296**, 1435.
- Ceponiene, R., Haapanen, M.-L., Ranta, R., Näätänen, R., & Hukki, J. (2002). Auditory sensory impairment in children with oral clefts as indexed by auditory event-related potentials. *Journal of Craniofacial Surgery*, **13**, 554–566.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, **1**, 351–353.
- Chomsky, N. (1980). *Rules and representations*. Oxford: Basil Blackwell.
- Cooper, R.P., & Aslin, R.N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, **65**, 1663–1677.
- Csibra, G., Davis, G., Spratling, M.W., & Johnson, M.H. (2000). Gamma oscillations and object processing in the infant brain. *Science*, **290**, 1582–1585.
- Csibra, G., & Johnson, M.H. (2007). Investigating event-related oscillations in infancy. In M. de Haan (Ed.), *Infant EEG and event-related potentials* (pp. 289–304). Hove: Psychology Press.
- Dale, A.M., & Sereno, M.I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction. *Journal of Cognitive Neuroscience*, **5**, 162–176.
- de Boer, B., & Kuhl, P.K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, **4**, 129–134.
- DeBoer, T., Scott, L.S., & Nelson, C.A. (2007). Methods for acquiring and analyzing infant event-related potentials. In M. de Haan (Ed.), *Infant EEG and event-related potentials* (pp. 5–37). Hove: Psychology Press.
- Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, **370**, 292–295.
- Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience*, **16**, 1375–1387.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 14240–14245.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, **134**, 9–21.
- Dien, J. (1998). Issues in the application of the average reference: review, critiques, and recommendations. *Behavior Research Methods, Instruments & Computers*, **30**, 34–43.

- Fellman, V., Kushnerenko, E., Mikkola, K., Ceponiene, R., Leipälä, J., & Näätänen, R. (2004). Atypical auditory event-related potentials in preterm infants during the first year of life: a possible sign of cognitive dysfunction? *Pediatric Research*, **56**, 291–297.
- Ferguson, C.A. (1964). Baby talk in six languages. *American Anthropologist*, **66**, 103–114.
- Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford: Oxford University Press.
- Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, **64**, 637–656.
- Guthrie, D., & Buchwald, J.S. (1991). Significance testing of difference potentials. *Psychophysiology*, **28**, 240–244.
- Hämäläinen, M.S., & Ilmoniemi, R.J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical and Biological Engineering and Computing*, **32**, 35–42.
- Hamburger, H.L., & van der Burgt, M.A. (1991). Global field power measurement versus classical method in the determination of the latency of evoked potential components. *Brain Topography*, **3**, 391–396.
- Hanson, H.M., & Stevens, K.N. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLSyn. *Journal of the Acoustical Society of America*, **112**, 1158–1182.
- Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, **97**, 3099–3111.
- Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, **47**, 359–382.
- Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neuroscience Research*, **54**, 276–280.
- Hwang, G., Jacobs, J., Geller, A., Danker, J., Sekuler, R., & Kahana, M.J. (2005). EEG correlates of verbal and nonverbal working memory. *Behavioral and Brain Functions*, **1**, 20.
- Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P.K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *NeuroReport*, **17**, 957–962.
- Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biology*, **6**, e11.
- Johnson, K., Flemming, E., & Wright, R. (1993). The hyper-space effect: phonetic targets are hyperarticulated. *Language*, **69**, 505–528.
- Johnson, M.H., de Haan, M., Oliver, A., Smith, W., Hatzakis, H., Tucker, L.A., & Csibra, G. (2001). Recording and analyzing high-density event-related potentials with infants using the Geodesic sensor net. *Developmental Neuropsychology*, **19**, 295–323.
- Kayser, J., & Tenke, C.E. (2006). Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. *Clinical Neurophysiology*, **117**, 348–368.
- Kirchhoff, K., & Schimmel, S. (2005). Statistical properties of infant-directed versus adult-directed speech: insights from speech recognition. *Journal of Acoustical Society of America*, **117**, 2238–2246.
- Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, **24**, 372–392.
- Klatt, D.H., & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, **87**, 820–857.
- Klimesch, W., Freunberger, R., Sauseng, P., & Gruber, W. (2008). A short review of slow phase synchronization and memory: evidence for control processes in different memory systems? *Brain Research*, **1235**, 31–44.
- Klimesch, W., Hanslmayr, S., Sauseng, P., Gruber, W., Brozinsky, C.J., Kroll, N.E., Yonelinas, A.P., & Doppelmayr, M. (2006). Oscillatory EEG correlates of episodic trace decay. *Cerebral Cortex*, **16**, 280–290.
- Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, **277**, 684–686.
- Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 979–1000.
- Kuhl, P.K., Tsao, F.M., & Liu, H.M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9096–9101.
- Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606–608.
- Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotila, M., & Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *NeuroReport*, **13**, 47–51.
- Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, **27**, 308–314.
- Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., & Fox, P.T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, **10**, 120–131.
- Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*, **48**, 609–621.
- Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., & Hämäläinen, M.S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage*, **31**, 160–171.
- Liu, H.M., Kuhl, P.K., & Tsao, F.M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, **6**, F1–F10.
- Liu, H.M., Tsao, F.M., & Kuhl, P.K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *Journal of Child Language*, **36**, 909–922.

- Love, T., Haist, F., Nicol, J., & Swinney, D. (2006). A functional neuroimaging investigation of the roles of structural complexity and task-demand during auditory sentence processing. *Cortex*, **42**, 577–590.
- Meltzoff, A.N., Kuhl, P.K., Movellan, J., & Sejnowski, T.J. (2009). Foundations for a new science of learning. *Science*, **325**, 284–288.
- Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D., & von Cramon, D.Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, **89**, 277–289.
- Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, **115**, 2195–2222.
- Mills, D.L., Prat, C., Zangl, R., Stager, C.L., Neville, H.J., & Werker, J.F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience*, **16**, 1452–1464.
- Molfese, D.L. (1989). The use of auditory evoked responses recorded from newborn infants to predict later language skills. *Birth Defects Original Article Series*, **25**, 47–62.
- Molfese, D.L., & Molfese, V.J. (1980). Cortical response of preterm infants to phonetic and nonphonetic speech stimuli. *Developmental Psychology*, **16**, 574–581.
- Morgan, J.L., & Demuth, K. (1996). Signal to syntax: an overview. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 1–22). Mahwah, NJ: Lawrence Erlbaum.
- 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**, 2544–2590.
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, **125**, 826–859.
- Novak, G.P., Kurtzberg, D., Kreuzer, J.A., & Vaughan, H.G. Jr. (1989). Cortical responses to speech sounds and their formants in normal infants: maturational sequence and spatiotemporal analysis. *Electroencephalography and Clinical Neurophysiology*, **73**, 295–305.
- Orehkova, E.V., Stroganova, T.A., Posikera, I.N., & Elam, M. (2006). EEG theta rhythm in infants and preschool children. *Clinical Neurophysiology*, **117**, 1047–1062.
- Perrin, F., Pernier, J., Bertrand, O., & Echallier, J.F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, **72**, 184–187.
- Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R. Jr., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., & Taylor, M.J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, **37**, 127–152.
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, **41**, 245–255.
- Polka, L., & Werker, J.F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 421–435.
- Radicevic, Z., Vujovic, M., Jelacic, L., & Sovilj, M. (2008). Comparative findings of voice and speech: language processing at an early ontogenetic age in quantitative EEG mapping. *Experimental Brain Research*, **184**, 529–532.
- Reynolds, G.D., & Richards, J.E. (2005). Familiarization, attention, and recognition memory in infancy: an event-related potential and cortical source localization study. *Developmental Psychology*, **41**, 598–615.
- Rivera-Gaxiola, M., Silva-Pereyra, J., Klarman, L., Garcia-Sierra, A., Lara-Ayala, L., Cadena-Salazar, C., & Kuhl, P.K. (2007). Principal component analyses and scalp distribution of the auditory P150-250 and N250-550 to speech contrasts in Mexican and American infants. *Developmental Neuropsychology*, **31**, 363–378.
- Saito, Y., Aoyama, S., Kondo, T., Fukumoto, R., Konishi, N., Nakamura, K., Kobayashi, M., & Tushima, T. (2007). Frontal cerebral blood flow change associated with infant-directed speech. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, **92**, F113–F116.
- Samar, V.J., Bopardikar, A., Rao, R., & Swartz, K. (1999). Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. *Brain and Language*, **66**, 7–60.
- 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**, 14–21.
- Scheeringa, R., Petersson, K.M., Oostenveld, R., Norris, D.G., Hagoort, P., & Bastiaansen, M.C. (2009). Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *NeuroImage*, **44**, 1224–1238.
- Schneider, T.R., Debener, S., Oostenveld, R., & Engel, A.K. (2008). Enhanced EEG gamma-band activity reflects multi-sensory semantic matching in visual-to-auditory object priming. *NeuroImage*, **42**, 1244–1254.
- Scott, S.K., & Wise, R.J. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, **92**, 13–45.
- Sharma, A., Marsh, C.M., & Dorman, M.F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *Journal of the Acoustical Society of America*, **108**, 3030–3035.
- Simos, P.G., Molfese, D.L., & Brenden, R.A. (1997). Behavioral and electrophysiological indices of voicing-cue discrimination: laterality patterns and development. *Brain and Language*, **57**, 122–150.
- Skipper, J.I., Nusbaum, H.C., & Small, S.L. (2005). Listening to talking faces: motor cortical activation during speech perception. *NeuroImage*, **25**, 76–89.
- Summerfield, C., & Mangels, J.A. (2005). Coherent theta-band EEG activity predicts item-context binding during encoding. *NeuroImage*, **24**, 692–703.
- Swingle, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 3617–3632.
- Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Reviews Neuroscience*, **5**, 721–728.
- Telkemeyer, S., Rossi, S., Koch, S.P., Nierhaus, T., Steinbrink, J., Poeppel, D., Obrig, H., & Wartenburger, I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience*, **29**, 14726–14733.

- Tonnquist-Uhlen, I. (1996). Topography of auditory evoked long-latency potentials in children with severe language impairment: the P2 and N2 components. *Ear and Hearing*, **17**, 314–326.
- Trainor, L.J., & Desjardins, R.N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, **9**, 335–340.
- Tremblay, K.L., Friesen, L., Martin, B.A., & Wright, R. (2003). Test–retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear and Hearing*, **24**, 225–232.
- Vallabha, G.K., McClelland, J.L., Pons, F., Werker, J.F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 13273–13278.
- Werker, J.F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, **103**, 147–162.
- Wilson, S.M., Saygin, A.P., Sereno, M.I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, **7**, 701–702.
- Woods, D.L., & Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology*, **65**, 447–459.
- Yao, D.Z., Wang, L., Oostenveld, R., Nielsen, K.D., Arendt-Nielsen, L., & Chen, A.C.N. (2005). A comparative study of different references for EEG spectral mapping: the issue of the neutral reference and the use of the infinity reference. *Physiological Measurement*, **26**, 173–184.
- Zangl, R., & Mills, D.L. (2007). Increased brain activity to infant-directed speech in 6- and 13-month-old infants. *Infancy*, **11**, 31–62.
- Zatorre, R.J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, **11**, 946–953.
- Zhang, Y., Kuhl, P.K., Imada, T., Iverson, P., Pruitt, J., Stevens, E.B., Kawakatsu, M., Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *NeuroImage*, **46**, 226–240.
- Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage*, **26**, 703–720.

Received: 29 September 2009

Accepted: 10 July 2010

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supplementary Figure 1 Schematic illustration of two stimulus presentation paradigms. (a) In the alternating block design, each block contains 20 stimuli of one category (A) followed by a block that contains stimuli of a different category (B). The blocks are sequentially alternated to collect sufficient trials for both stimuli. (b) In the bidirectional MMN paradigm, two long blocks are presented to represent change directions (i.e. from A to B or B to A). The MMN response to stimulus change is derived by subtracting the brain responses to the same stimuli to control for inherent acoustic differences between A and B (i.e. deviant A minus standard A and deviant B minus standard B). In theory, the bidirectional MMN paradigm allows the study of neural representations of the auditory stimuli as well as sensitivity of the neural system to change detection (Zhang et al., 2005). However, the time length that is required to collect sufficient trials in the bidirectional MMN design would exceed the limit for most infant participants at 6–12 months of age.

Supplementary Sounds 1 and 2 The sound1.wav file was the non-exaggerated /i/. The sound2.wav file was the formant-exaggerated /i/. A preliminary survey on 15 adult native English speakers (all were naïve listeners who had not been previously tested on synthetic speech in laboratory settings) indicated that both sounds were heard as /i/s. Twelve listeners reported the /i/ with expanded formants to be clearer or perceptually more exaggerated than the other /i/ sound. The other three did not give unequivocal judgment about which sound was perceptually clearer or more exaggerated, which was probably due to unfamiliarity with synthetic speech.

Supplementary Movies 1 and 2 The MNE (minimum norm estimate) movie files were generated in BESA software and converted to compressed format. The movie1.wmv file showed left-hemisphere MNE activities for the exaggerated /i/, which were scaled to the maximum MNE peak in percentages and color-coded. The movie2.wmv file showed right-hemisphere MNE activities for the exaggerated /i/ using the peak-scaled color coding scheme. A cautionary note is necessary here. As the peak MMN values for different stimuli were not the same, a direct comparison of the color-coded brain activation patterns could lead to misinterpretation.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.