

Voice and Emotion Processing in the Human Neonatal Brain

Yawei Cheng^{1,2}, Shin-Yi Lee^{1*}, Hsin-Yu Chen^{1*},
Ping-Yao Wang², and Jean Decety³

Abstract

■ Although the voice-sensitive neural system emerges very early in development, it has yet to be demonstrated whether the neonatal brain is sensitive to voice perception. We measured the EEG mismatch response (MMR) elicited by emotionally spoken syllables “dada” along with correspondingly synthesized nonvocal sounds, whose fundamental frequency contours were matched, in 98 full-term newborns aged 1–5 days. In Experiment 1, happy syllables relative to nonvocal sounds elicited an MMR lateralized to the right hemisphere.

In Experiment 2, fearful syllables elicited stronger amplitudes than happy or neutral syllables, and this response had no sex differences. In Experiment 3, angry versus happy syllables elicited an MMR, although their corresponding nonvocal sounds did not. Here, we show that affective discrimination is selectively driven by voice processing per se rather than low-level acoustical features and that the cerebral specialization for human voice and emotion processing emerges over the right hemisphere during the first days of life. ■

INTRODUCTION

Voice communication, at the heart of human life, is critical for survival and social communication (Grossmann & Friederici, 2011; Belin & Grosbras, 2010; Belin, Fecteau, & Bedard, 2004). The voice carries the acoustical signature of our species, which conveys important affective and identity information (Latinus & Belin, 2011; Belin et al., 2004). In human adults, voices are specifically processed in the upper bank of the STS (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). In macaque monkeys, voices are processed in the superior temporal plane that preferentially responds to conspecific vocalizations. Recognizing the vocalization of a species member is independent of language, and is an evolutionarily conserved brain function (Belin & Grosbras, 2010; Petkov, Logothetis, & Obleser, 2009; Petkov et al., 2008).

The temporal voice areas (TVAs) show greater activity in response to voices (speech and nonspeech vocalizations such as laughs, cough, etc.) than to natural nonvocal sounds (environmental sounds, musical sounds, animal vocalizations, etc.), or amplitude- or frequency-matched acoustical control sounds (Kriegstein & Giraud, 2004; Belin et al., 2000). Of note, the right TVA shows strong sensitivity to affective information crucial in social communication (Belin & Grosbras, 2010; Ethofer et al., 2006; Grandjean et al., 2005). Such sensitivity is particularly

strong for threat-related emotions (e.g., fear and anger), which are processed independently of attention and considered as a fundamental neural mechanism that prioritizes the processing of social stimuli. (Decety, 2011; Vuilleumier, 2005; Belin et al., 2004).

Interestingly, voice perception abilities seem to appear earlier than speech perception in human development. Although phoneme discrimination emerges in 2-month-olds and lexical-semantic processing, in 12- to 14-month-olds (Friederici, 2005; Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002), infants already exhibit well-developed abilities for voice perception. Behavioral studies have demonstrated that newborn infants prefer human voices to nonvocal auditory stimuli (Ecklund-Flores & Turkewitz, 1996; Hutt, von Bernuth, Lenard, Hutt, & Prechtel, 1968) and their native language to a foreign language with different prosodic characteristics (Moon, Cooper, & Fifer, 1993; Mehler et al., 1988). Behavioral and electrophysiological evidence supports that newborns can discriminate their mother's voice from the voice of another woman (Beauchemin et al., 2011; Ockleford, Vince, Layton, & Reader, 1988; DeCasper & Fifer, 1980). When presented with vocal expressions with a range of emotional prosodies (happy, angry, sad, and neutral), newborns showed an increase in eye-opening responses following happy prosody compared with the other emotional expressions but only while listening to their maternal language (Mastropieri & Turkewitz, 1999). Importantly, this voice sensitivity is even present in fetuses before birth (Kisilevsky et al., 2003). Newborns have extracted a number of prosodic (rhythmic and intonational) characteristics of auditory input during the last few days or weeks of

¹Institute of Neuroscience and Brain Research Center, National Yang-Ming University, Taipei, Taiwan, ²National Yang-Ming University Hospital, Yilan, Taiwan, ³The University of Chicago

*These authors contributed equally to this work.

prenatal development. In spite of this very early form of sensitivity to happy prosody in familiar contexts, a number of neuroimaging studies indicated that only from around 3 to 7 months of age do infants could process the voice and its emotions (Blasi et al., 2011; Lloyd-Fox, Blasi, Mercure, Elwell, & Johnson, 2011; Grossmann, Oberecker, Koch, & Friederici, 2010; Flom & Bahrick, 2007; Walker-Andrews, 1997). Here, we report three electrophysiological experiments with newborns designed to examine the neonatal brain sensitivity to voice perception.

Recently, the emergence of TVA was reported in 3- to 7-month-old infants (Blasi et al., 2011; Lloyd-Fox et al., 2011; Grossmann, 2010). However, it is possible that the voice sensitivity in the 7-month-old brain reflects the processing of low-level acoustical features in vocal sounds rather than in voice processing per se. Particularly, the greater response to affective (happy and angry) compared with neutral voices could reflect the processing of the many acoustical differences between these sound categories without implying emotional processing (Belin & Grosbras, 2010). Fundamental frequency, f_0 , is considered the most correlative acoustical variable of emotions, which decreases over time for angry prosody but increases for happy prosody (Banse & Scherer, 1996). Thus, in the current study, we used another set of control sounds, in which nonvocal sounds were created to follow the same f_0 contours as vocal sounds, to test the influences of low-level acoustical structure.

Here, in full-term newborns, we measured the mismatch response (MMR), an infant equivalent of adult MMN, in response to emotionally spoken syllables “dada,” which is known to elicit emotional MMN in adults (Schirmer, Striano, & Friederici, 2005). MMR may occur as a positive or negative deflection in infants (Csibra, Kushnerenko, & Grossmann, 2008; Maurer, Bucher, Brem, & Brandeis, 2003; Cheour, Kushnerenko, Čeponienė, Fellman, & Näätänen, 2002; Friederici, Friedrich, & Weber, 2002). MMR has been used to demonstrate the emergence of sound and speech perception in newborns (Čeponienė et al., 2002; Cheour-Luhtanen et al., 1995). In addition, MMN, a component of ERPs, is a reliable indicator for automatic (i.e., attention-independent) auditory discrimination (Näätänen, Paavilainen, Rinne, & Alho, 2007). The MMN has also been shown to reflect the affective discrimination of voice perception in adults (Schirmer et al., 2005). Hearing emotional relative to neutral syllables elicited a stronger MMN in women, not men, which might be associated with sex differences in this aspect of social orientation (Schirmer et al., 2008). Furthermore, the MMN may be generated by neural sources in primary and nonprimary auditory areas in the superior temporal cortex, including those known to be selectively involved in voice processing (Herrmann, Maess, Hasting, & Friederici, 2009; Näätänen et al., 2007).

We first investigated the voice sensitivity in newborns, as behavioral studies have shown voices to be processed at birth. To examine whether newborns are already sensitive to human voice and able to discriminate vocal from non-

vocal sounds, Experiment 1 measured the MMR to happily spoken syllables “dada” as the deviant and the corresponding nonvocal sound as the standard in an oddball paradigm. In the second experiment, we assessed whether emotional prosodies modulate the ability for voice sensitivity identified in Experiment 1 (Belin & Grosbras, 2010; Ethofer et al., 2006; Grandjean et al., 2005), and whether this modulation, if present, differs between sexes (Schirmer et al., 2005, 2008). Newborns were presented with happy, fearful, and neutral syllables while their MMR was recorded. Finally, the third experiment was conducted to further determine whether this affective discrimination is driven by low-level acoustical parameters or not (Belin & Grosbras, 2010; Belin et al., 2000). In this latter experiment, newborns were tested with happy and angry syllables as well as corresponding nonvocal sounds. The nonvocal sounds were synthesized to follow the f_0 contours of emotional syllables.

METHODS

Participants

The study sample consisted of 25 newborns (10 girls) in Experiment 1, 43 newborns (20 girls) in Experiment 2, and 30 newborns (11 girls) in Experiment 3. An additional 15 newborns were tested ($n = 4$ in Experiment 1, $n = 6$ in Experiment 2, and $n = 5$ in Experiment 3) but were not included in the final sample because of motion artifacts resulting in too few usable trials for data analysis (minimal number of 60 trials per condition; $n = 12$) or technical failure ($n = 3$). All neonates aged between 0 and 5 days ($M = 2.6, 1.6,$ and 1.5 days for Experiments 1–3, respectively) were born full-term (37–42 weeks gestation) and with normal birth weight (2595–3890 g). They passed a hearing screening with evoked otoacoustic emissions and were declared healthy by neonatologists. All of their parents gave informed consent before the study. The study was approved by the local ethics committee of the National Yang-Ming University Hospital.

Auditory Stimuli

For Experiment 1, the stimulus material consisted of the happily spoken syllables “dada” and its corresponding nonvocal sounds. For Experiment 2, the stimulus material consisted of fearful, happy, and neutral syllables. For Experiment 3, the stimulus material consisted of happy and angry syllables as well as their corresponding nonvocal sounds.

A young female speaker (25 years old) produced the syllables of “dada” with three sets of emotional prosodies (fearful, angry, and happy) and one set of neutral prosody. Within a kind of emotional or neutral prosody, the speaker produced the syllables “dada” for more than 10 times. Syllables were edited to become equally long (550 msec) and loud (min: 57 dB, max: 62 dB; mean: 59 dB) with the use of Cool Edit Pro 2.0 and Sound Forge 9.0. Stimuli were rated

for emotionality by 120 listeners (60 men). For the fearful set, listeners classified each stimulus with a 5-point scale from *extremely fearful* to *not fearful at all*. For the angry set, listeners classified each stimulus with a 5-point scale from *extremely angry* to *not angry at all*. For the happy set, listeners classified each stimulus with a 5-point scale from *extremely happy* to *not happy at all*. Three emotional syllables that had been consistently identified as *extremely fearful*, *extremely angry*, and *extremely happy* were selected as the experimental stimuli. The neutral syllables rated as the most emotionless were selected as the control stimulus.

Furthermore, the corresponding set of four nonvocal sounds that follow the envelope of each emotional (fearful, angry, happy, and neutral) syllables were created by Praat (Boersma, 2001). For each original syllable, a sine waveform was synthesized at a modulated frequency following the original f_0 contour. All of the stimuli did not differ with respect to their duration and mean intensity.

Procedures

Neonates were tested while lying on a bed in a sound-attenuated room of the hospital. Stimuli were presented via two loudspeakers placed at approximately 25 cm on the right and left sides of the newborn's head. A mean background noise level was 35-dB SPL (sound pressure level).

Experiment 1 comprised two or three blocks. Each block lasted 4 min, containing 200 stimuli in an auditory oddball paradigm with a 1200-msec sound onset asynchrony. Every block consisted of one deviant ($p = .2$) and one standard stimuli ($p = .8$). The happy syllables were set as the deviant, and its corresponding nonvocal sounds as the standard. The deviants ran at a random order of sequences, edited by Matlab 7.0 (MathWorks, Natick, MA). Each deviant followed at least two standards. In Experiment 2, there were four to six blocks. Every block consisted of one standard ($p = .8$) and two deviant stimuli ($p = .1$). The neutral syllables were set as the standard, and the emotional syllables were the two deviants (Deviant 1: happy, Deviant 2: fear). In Experiment 3, there was one session for emotional syllables and another session for their corresponding nonvocal sounds. Each session following the same auditory oddball paradigm as in Experiment 1 included the stimuli belonging to the same category (emotional syllables or nonvocal sounds). The happy was set as the standard, and the angry was the deviant. The order of the sessions was randomized across participants. The experimental duration, including preparation and break, never exceeded 1 hr. If newborns became hungry or started crying, the experimental procedure was stopped.

EEG Apparatus and Recording

EEG data were collected from eight single-used Ag/Ag-Cl electrodes at F3, F4, C3, C4, T3, T4, P3, and P4 according to

the International 10–20 System. The reference electrodes were placed over the right mastoid (A2), and the ground electrode was on the forehead. The EOG-monitored eye blinks and vertical eye movements from the electrodes positioned above and below the right eye, and recorded horizontal eye movements from the electrodes placed laterally to the right and left external canthi. Electrolytic gel was applied at each electrode to reduce the impedance of the electrode–skin contact. The impedance of all electrodes was maintained below 5 Ω .

Data were recorded using the NuAmp amplifier and analyzed with Neuroscan 4.3 (Compumedics Ltd., Victoria, Australia). Channels were rereferenced off-line to the average of the left and right mastoids $[(A1 + A2) / 2]$, sampled at 250 Hz, and band-pass filtered (0.1–30 Hz). Trials were epoched over an analysis time of 900 msec, including pre-stimulus of 100 msec for baseline correction. Any epoch contaminated by eye blinks, eye movements, or muscle potentials exceeding $\pm 150 \mu\text{V}$ at any electrode was automatically excluded from the average. The signal quality was further ensured by careful visual inspection in every participant and trial and application of a digital 1–15 Hz zero-phase band-pass filter with a 12-dB/octave slope. To eliminate the possibility of attention transients because of the stimulus onset, the first three epochs of each block were excluded from data analysis. Because previous studies with newborns did not find differences in MMR latencies and amplitudes and distribution between sleep and awaken stages (e.g., Hirasawa, Kurihara, & Konishi, 2002), the signal was recorded for analysis whenever neonates were either awake or asleep during the experiments.

The dependent variable was the mean ERP amplitudes, obtained from each grand-averaged peak (± 50 msec) within a time window of 300–500 msec after stimulus onset. With the use of Matlab 7.0 and SPSS 18.0 (SPSS, Inc., Chicago, IL), the moving time-window technique (Luck, 2005) determined the target window for each experiment. Considering the frontocentral distribution for MMR (Näätänen et al., 2007) and unclosed fontanel in newborns, only a subset of the recorded electrodes (F3, F4, C3, and C4) were included for analysis. For Experiment 1, a three-way repeated-measures ANOVA (Stimulus [happy syllables vs. nonvocal sounds] \times Lateralization [left vs. right hemispheres] \times Region [frontal vs. central electrodes]) was conducted. For Experiment 2, statistical analysis was examined through a four-way mixed ANOVA with three within-subject factors (Stimulus [neutral vs. happy vs. fearful] \times Lateralization [left vs. right hemispheres] \times Region [frontal vs. central electrodes]). Subtracting the neutral ERP from the emotional ERP produced the corresponding MMR. For Experiment 3, a four-way repeated-measures ANOVA (Session [emotional syllables vs. nonvocal sounds] \times Stimulus [happy vs. angry] \times Lateralization [left vs. right hemispheres] \times Region [frontal vs. central electrodes]) was computed. Subtracting the happy ERP from the angry ERP produced the angry–happy MMR. Degrees of freedom were corrected using the Greenhouse–Geisser method.

Bonferroni–Dunn test was conducted only when preceded by significant effects.

RESULTS

For Experiment 1, mean ERP amplitudes at 400–500 msec showed main effects of Stimulus [$F(1, 24) = 5.28, p = .031, d = 0.94$] and Lateralization [$F(1, 24) = 4.65, p = .041, d = 0.88$] as well as their interaction [Stimulus \times Lateralization: $F(1, 24) = 7.07, p = .014, d = 1.09$]. The happy syllables (deviant) elicited significantly stronger amplitudes than the happy-derived nonvocal sounds (standard). The right hemisphere showed stronger responses than the left hemisphere. Post hoc analysis found the stimulus effect over the right hemisphere [$t(49) = -3.81, p < .001, d = -0.54$] but none over the left hemisphere [$t(49) = -1.01, p = .316, d = -0.14$]. The MMR to discriminate the happy syllables from the nonvocal sounds occurred at F4 [$t(24) = 2.74, p = .011, d = 0.55$] and C4 [$t(24) = 2.60, p = .016, d = 0.52$] but not at F3 [$t(24) = 0.45, p = .655, d = 0.09$] and C3 [$t(24) = 0.97, p = .344, d = 0.19$; Figure 1].

For Experiment 2, ERP at the time window of 350–450 msec reached significance for Stimulus [$F(2, 82) = 4.93, p = .010, d = 0.69$] and Region [$F(1, 41) = 16.87, p < .001, d = 1.28$] as well as their interaction [$F(2, 82) = 6.82, p = .002, d = 0.82$]. However, there was no effect in Sex [$F(1, 41) = 0.51, p = .479, d = 0.22$] and Lateralization [$F(1, 41) = 0.10, p = .757, d = 0.09$]. Follow-up

analysis revealed significant differences among emotional (fearful vs. happy vs. neutral) syllables at electrodes F3 [$F(2, 82) = 3.96, p = .023, d = 0.62$], F4 [$F(2, 82) = 10.24, p < .001, d = 1.00$], and C3 [$F(2, 82) = 3.63, p = .031, d = 0.59$] but not at C4 [$F(2, 82) = 1.05, p = .354, d = 0.32$]. Post hoc analysis indicated that fearful (Deviant 2) relative to neutral (standard) syllables elicited stronger positivity at F3 [$t(42) = 3.38, p = .002, d = 0.57$] and F4 [$t(42) = 4.52, p < .001, d = 0.77$]. The happy syllables (Deviant 1) differed from the neutral syllables (standard) at F4 [$t(42) = 2.12, p = .040, d = 0.34$], not at F3 [$t(42) = 1.10, p = .279, d = 0.21$]. Significant differences between fearful and happy syllables occurred at F4 over the right hemisphere [$t(42) = 2.39, p = .021, d = 0.44$], not at F3 over the left hemisphere [$t(42) = 1.45, p = .154, d = 0.27$; Figure 2].

The ERP at the time window of 550–650 msec of Experiment 2 reached significance for the Stimulus [$F(2, 82) = 3.42, p = .038, d = 0.58$] and for the interaction of Lateralization \times Region [$F(1, 41) = 4.77, p = .035, d = 0.68$]. However, there was no effect of Region [$F(1, 41) = 0.04, p = .842, d = 0.06$], Lateralization [$F(1, 41) = 0.13, p = .726, d = 0.11$], and Sex [$F(1, 41) = 0.74, p = .396, d = 0.27$]. Follow-up analysis revealed significant differences among emotional (fearful vs. happy vs. neutral) syllables at electrodes F4 [$F(2, 82) = 5.96, p = .004, d = 0.26$] and C4 [$F(2, 82) = 3.70, p = .029, d = 0.60$] but not at F3 [$F(2, 82) = 1.88, p = .160, d = 0.43$] and C3 [$F(2, 82) = 0.37, p = .696, d = 0.19$]. Post hoc analysis showed that the fearful (deviant) relative to neutral (standard) syllables elicited stronger positivity at F4 [$t(42) = 2.07, p = .045,$

Figure 1. Event-related brain potentials to hearing happily spoken syllables “dada” relative to the synthesized nonvocal sounds in newborns ($n = 25$). (A) Oscillogram of happy “dada” and corresponding nonvocal sound. Their mean intensities are matched (76.37 dB vs. 76.32 dB). (B) Spectrogram of happy “dada” and corresponding nonvocal sound. The nonvocal sound was synthesized to follow the f_0 contour of happy syllables. (C) The deviant (happy syllables) significantly differs from the standard (nonvocal sounds) at electrodes F4 and C4 ($p = .011$ and $p = .016$, black arrow) over the right hemisphere but not at F3 and C3 ($p > .05$ and $p > .05$, white arrow) over the left hemisphere.

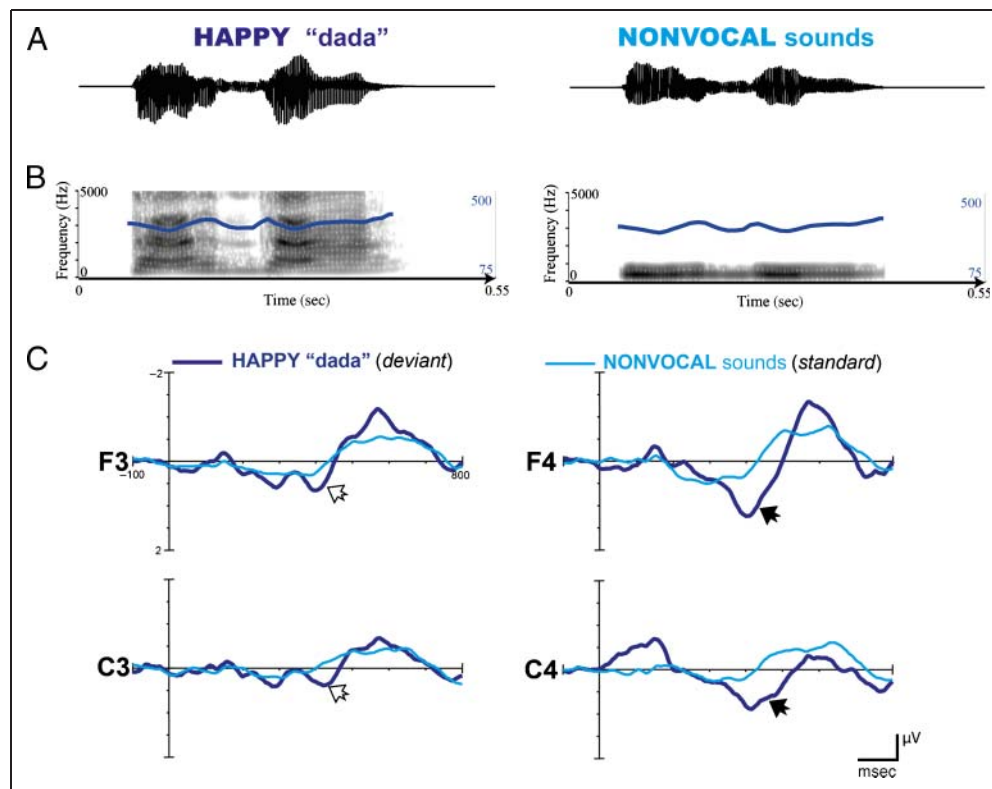
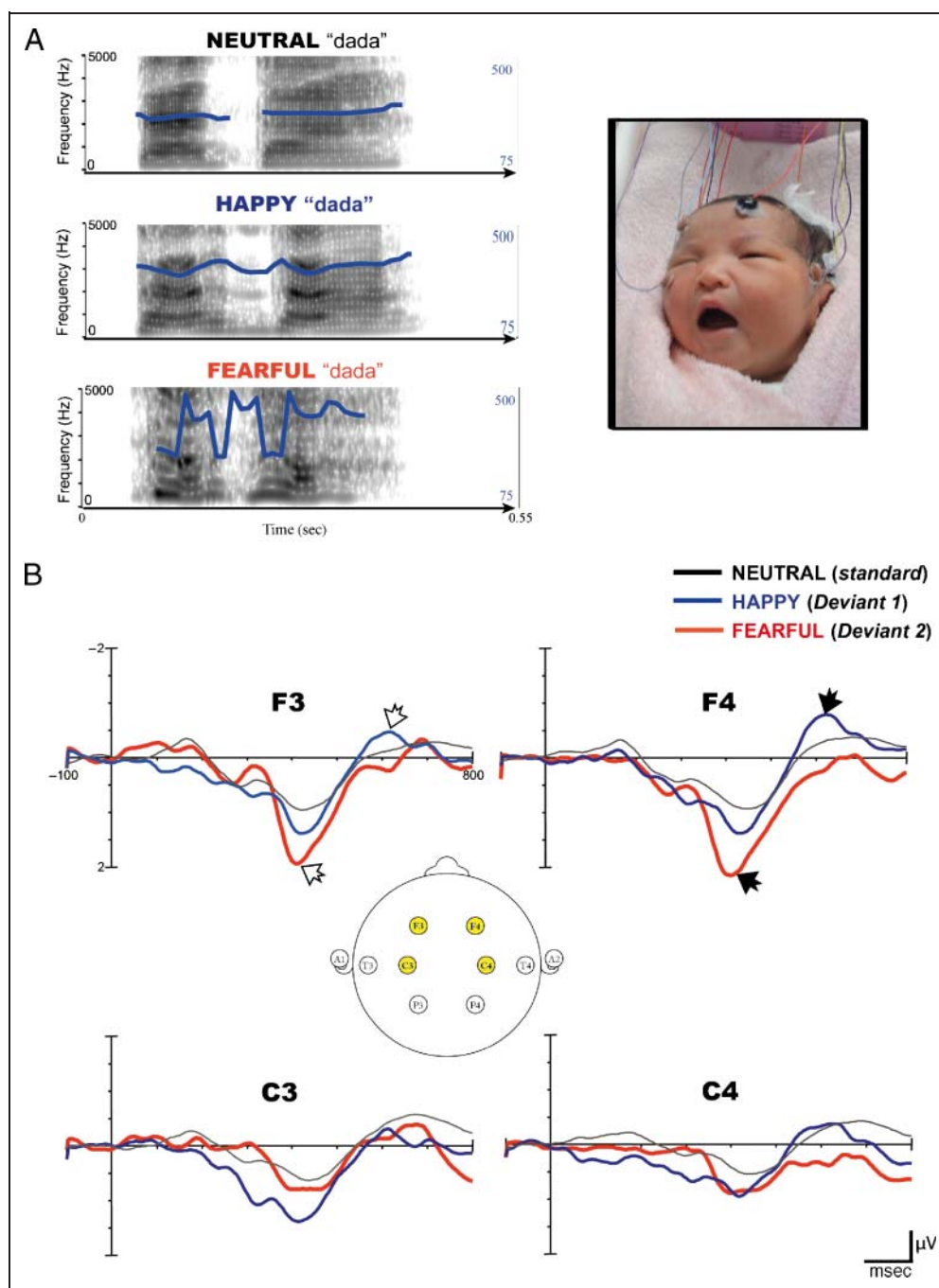


Figure 2. Event-related brain potentials to hearing emotionally spoken syllables “dada” in newborns ($n = 43$). (A) Spectrogram of the “dada” syllables produced with neutral, happy, and fearful prosodies. (B) Event-related brain potentials to emotional syllables. There was a stimulus effect (neutral vs. happy vs. fearful) at the time window of 350–450 ($p = .010$) and 550–650 ($p = .038$) msec. The fearful relative to happy syllables elicited stronger positivity at electrode F4 over the right hemisphere ($p = .021$ and $p = .006$, black arrow), not at F3 over the left hemisphere ($p > .05$ and $p > .05$, white arrow).



$d = 0.42$] and C4 [$t(42) = 2.20, p = .034, d = 0.47$]. Significant differences between fearful and happy syllables occurred at F4 over the right hemisphere [$t(42) = 2.88, p = .006, d = 0.62$] but not at F3 over the left hemisphere [$t(42) = 1.57, p = .125, d = 0.36$].

For Experiment 3, ERPs at the time window of 300–400 msec reached significance for Stimulus [$F(1, 29) = 8.22, p = .008, d = 1.07$] and Region [$F(1, 29) = 5.99, p = .002, d = 1.28$] but none for Session [$F(1, 29) = 0.07, p = .790, d = 0.09$] and Lateralization [$F(1, 29) = 1.76, p = .195, d = 0.49$]. The angry syllables (deviant) elicited stronger amplitudes than the happy syllables (standard). The frontal electrodes showed stronger responses than the

central electrodes. With regard to the marginal interaction of Session \times Lateralization \times Region [$F(1, 29) = 3.97, p = .056, d = 0.74$], post hoc analysis found that the stimulus effect was mainly driven by the session for emotional syllables [$F(1, 29) = 8.79, p = .006, d = 1.10$] but not by the session for nonvocal sounds [$F(1, 29) = 1.67, p = .207, d = 0.48$]. Both sessions had no lateralization effect [$F(1, 29) = 3.65, p = .066, d = 0.71; F(1, 29) < 0.01, p = .955, d < 0.01$]. Within the session for emotional syllables, the MMR discrimination of the angry syllables from the happy syllables significantly occurred at F3 [$t(29) = 3.49, p = .002, d = 0.64$] and F4 [$t(29) = 2.40, p = .023, d = 0.44$] and marginally at C3 [$t(29) = 2.05, p = .050$,

$d = 0.37$] but not at C4 [$t(29) = 1.22, p = .234, d = 0.22$; Figure 3].

DISCUSSION

This study clearly demonstrates neural processing of voice sensitivity and affective prosody discrimination during the first days of life. For the first time, MMR points out that humans at birth possess voice sensitivity abilities. The finding indicates that the emergence and maturation of cerebral specialization for human voice emerges in the first five days of life, enabling newborns to be socially responsive.

The neonatal MMR to spoken syllables “dada” taps into the developmental origins of voice processing in the human brain. In Experiment 1, the presence of neonatal MMR for happy syllables relative to corresponding nonvocal sounds suggests that the capacity for processing human voices emerges very early during the neonatal period. One study reported the emergence of voice sensitivity, as indicated by stronger responses in the right TVA to the human vocal relative to nonvocal sounds, in 7-month-olds but not in

4-month-olds (Grossmann et al., 2010). This discrepancy may result from a number of differences in stimulus content, experimental paradigms (oddball vs. block), and neuroimaging techniques (ERP vs. functional near-infrared spectroscopy). First, the vocal stimuli consisted of non-speech (i.e., crying and laughing) as well as speech (words and nonwords) stimuli. It is difficult to ascertain whether the differential response across two age groups was related to speech perception or to the processing of other types of vocalization. Second, the nonvocal stimuli included animal vocal sounds, which may share features with those emitted by humans. It is unknown whether this had an impact on the strength of the contrast between the vocal and nonvocal conditions. Third, the nonvocal stimuli contained a mixture of sounds with heterogeneous familiarity to infants. It is undetermined whether degree in stimulus familiarity had an effect on the contrasts. Finally, lacking the control sounds designed to acoustically match human vocal sounds makes it difficult to ascertain whether the brain response for voice sensitivity reflects the processing of low-level acoustical features or voice-specific processing per se (Belin & Grosbras, 2010; Belin et al., 2000). Although

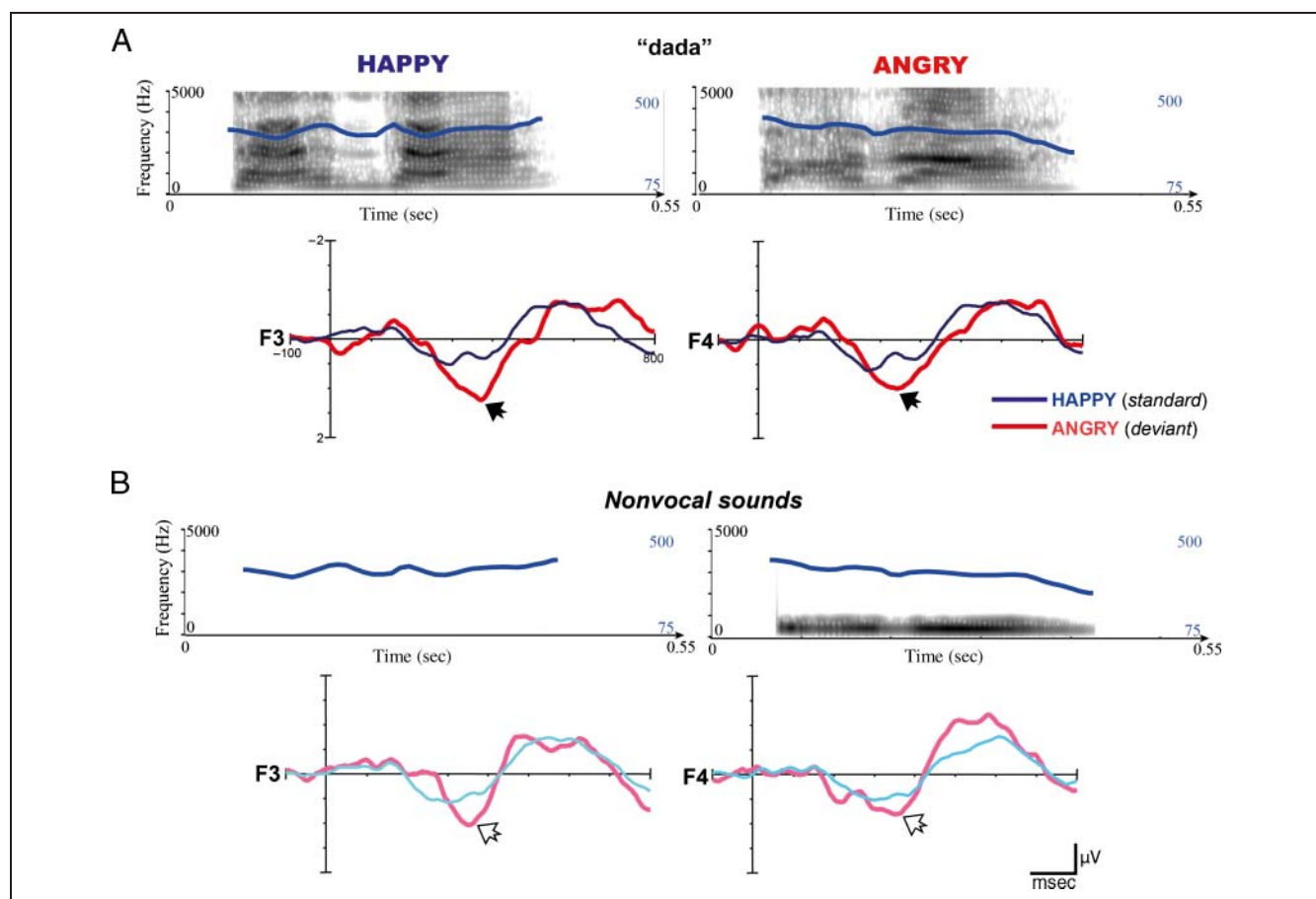


Figure 3. MMR to hearing emotionally spoken syllables “dada” and their corresponding nonvocal sounds, respectively, in newborns ($n = 30$). (A) Spectrogram indicates that happy syllables “dada” increase f_0 over time, but angry syllables decrease f_0 over time. The angry relative to happy syllables elicit significant MMR at electrodes F3 and F4 ($p = .002$ and $p = .023$, black arrow). (B) Spectrogram indicates that the nonvocal sounds are synthesized to follow the f_0 contours of happy and angry syllables, respectively. There is no significant MMR between their corresponding nonvocal sounds ($p > .05$ and $p > .05$, white arrow).

considering f_0 and intensity as the most correlative acoustical determinants of emotions (Banse & Scherer, 1996), we synthesized the corresponding nonvocal sounds that follow the envelope of each emotional (fearful, angry, happy, and neutral) syllables by Praat (Boersma, 2001) with the control of their duration and mean intensity (see Figure 1).

Furthermore, the presence of neonatal MMR for happy syllables relative to corresponding nonvocal sounds indicates that the TVA specialization in processing voices emerges very early in development. This may not be surprising as the primary auditory cortex, involved in voice processing, undergoes intensive synaptogenesis between 27 weeks and 3 months postterm (Huttenlocher & Dabholkar, 1997). This critical period has been shown to parallel that of dendritic development and myelination in the auditory cortex (Huttenlocher & Dabholkar, 1997; Barkovich, 1990). Interestingly, one recent study measured the electrophysiological index on the auditory oddball paradigm and demonstrated that newborns can distinctly process their mother's voice at an early preattentive level (Beauchemin et al., 2011). In parallel with the fronto-temporal positivity to voices in adults (Charest et al., 2009) and children (Rogier, Roux, Belin, Bonnet-Brilhault, & Bruneau, 2010), the fronto-central MMR to the discrimination between spoken syllables and nonvocal sounds appears as positive deflections in newborns. Thus, the neural underpinnings of human voice processing are already present at birth. MMR, as a biological indicator of voice sensitivity, may be crucial in assessing such cortical function in newborns, especially of those infants at risk for neurodevelopmental disorders with social reciprocity problems, such as autism spectrum disorders.

Our findings also indicate that voice affect contributes to the neonatal MMR. In Experiment 3, the angry-happy MMR was present for emotional syllables but absent for acoustical controls. The emergence of neonatal MMR to the perception of emotional syllables, being mainly driven by affective discrimination beyond acoustical distinction, is in good agreement with the available behavioral literature. For instance, one study examined neonates' eye widening in reaction to speech produced in either the mothers' native language or a novel language using neutral, happy, sad, or angry prosody (Mastropieri & Turkewitz, 1999). Increased eye widening was found in response to happy prosody but only for speech produced by the maternal language. This early discrimination of voice emotion seems to be related to prenatal experience with a specific language. Our study measured the neonatal MMR in reaction to emotional syllables "dada," irrelevant to semantics, which can reflect vocal affective information beyond specific language.

Negativity bias emerges early in the neonatal period. Here, in Experiments 2 and 3, newborns showed a stronger MMR, independent of attention, when hearing fearful and angry syllables relative to happy syllables. MMR represents attention-independent perceptual change detection (Näätänen et al., 2007). Angry prosody elicited a more

negative-going ERP and a stronger TVA activation than happy or neutral prosody in 7-month-olds (Grossmann, 2010; Grossmann, Striano, & Friederici, 2005). Angry syllables evoked a stronger MMN than happy or neutral syllables in adults (Schirmer et al., 2005). From an evolutionary perspective, the processing of threat-related emotions (e.g., fear and anger) is particularly strong, independent of attention (Belin & Grosbras, 2010; Vuilleumier, 2005).

Furthermore, the affective voice processing in newborns is characterized by a right hemispheric dominance. In spite of prominent MMR obtained from newborns in response to complex harmonic tones and speech sounds (Cheour, Leppanen, & Kraus, 2000), such response has no hemispheric asymmetry. In our study, the MMR in reaction to happy syllables versus nonvocal sounds (Experiment 1) as well as the differential MMR for fearful versus happy syllables (Experiment 2) were lateralized to the right hemisphere. Previous work has demonstrated stronger responses in the right superior temporal cortex and the right inferior frontal cortex to the emotional compared with neutral prosodies in 3- to 7-month-olds (Blasi et al., 2011; Grossmann et al., 2010). Voice-sensitive activity, which conveys affective cues, appeared stronger in the right than left TVA (Latinus & Belin, 2011; Belin et al., 2000). These results are consistent with the view that emotion processing is characterized by a right hemispheric dominance (Fox, 1991). The right TVA shows strong sensitivity to affective information crucial for social communication (Belin & Grosbras, 2010; Ethofer et al., 2006; Grandjean et al., 2005).

Behavioral studies have shown that newborns are capable of responding to cries emanating from other neonates (Dondi, Simion, & Caltran, 1999; Martin & Clark, 1982). This early affective arousal and discrimination is thought to play a crucial role as a building block of empathy (Decety, 2010). Neonates appear to possess the neural mechanism to discriminate emotions from voices. Emotions differentially modulate voice processing in the right hemisphere already in neonatal period.

In Experiment 2, we found that the MMR, as an indicator for affective discrimination, does not differ between genders. However, in young adults, one study reported that MMN to emotional syllables with the same auditory oddball paradigm exhibited gender differences (Schirmer et al., 2005). This finding suggests that neonatal brain may use different (immature) brain mechanisms to process this discrimination. Sexual dimorphism of voice sensitivity may be a result of sociocultural factors and sex hormones during adolescence.

Finally, the findings of our study have important implications for autism, a neurodevelopmental disorder. Adults with autism fail to activate voice-sensitive regions in temporal cortex (Gervais et al., 2004). Children and adults with autism have difficulties to recognize emotional expressions through voice (Rutherford, Baron-Cohen, & Wheelwright, 2002; Hobson, Ouston, & Lee, 1989; Van Lancker, Cornelius, & Kreiman, 1989). On the basis of a

relatively large sample ($n = 98$), our findings robustly demonstrated that MMR for affective discrimination is already specialized during the neonatal period. Amplitude of MMN in response to pure tone is associated with the functioning status of autism (Dunn, Gomes, & Gravel, 2008; Ferri et al., 2003). Therefore, in future work, the current approach could be used to assess individual differences in infants' MMR to affective voices and might thus serve as one of potentially multiple markers that can help with early identification of infants at risk for autism (Belin & Grosbras, 2010; Elsabbagh & Johnson, 2007).

Acknowledgments

The study was sponsored by the National Science Council (NSC 99-2314-B-010-037-MY3; NSC 100-2628-H-010-001-MY3), the National Yang-Ming University Hospital (RD2010-004; RD2011-005), the Academia Sinica (AS-99-TP-AC1), and a grant from the Ministry of Education (Aim for the Top University Plan). Dr. Jean Decety was supported by an NSF grant (BCS-0718480).

Reprint requests should be sent to Dr. Yawei Cheng, Institute of Neuroscience and Brain Research Center, National Yang-Ming University, 155, Sec. 2, St. Linong, Dist. Beitou, Taipei 112, Taiwan, R.O.C., or via e-mail: ywcheng2@ym.edu.tw.

REFERENCES

- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636.
- Barkovich, A. (1990). Normal development of the neonatal and infant brain. In A. Barkovich (Ed.), *Pediatric neuroimaging* (pp. 5–34). New York: Raven Press.
- Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martinez-Montes, E., Belin, P., et al. (2011). Mother and stranger: An electrophysiological study of voice processing in newborns. *Cerebral Cortex*, *21*, 1705–1711.
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Science*, *8*, 129–135.
- Belin, P., & Grosbras, M. H. (2010). Before speech: Cerebral voice processing in infants. *Neuron*, *65*, 733–735.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.
- Blasi, A., Mercure, E., Lloyd-Fox, S., Thoms, A., Brammer, M., Sauter, D., et al. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, *21*, 1220–1224.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, *5*, 341–345.
- Čeponienė, R., Kushnerenko, E., Fellman, V., Renlund, M., Suominen, K., & Näätänen, R. (2002). Event-related potential features indexing central auditory discrimination by newborns. *Brain Research, Cognitive Brain Research*, *13*, 101–113.
- Charest, I., Pernet, C. R., Rousselet, G. A., Quinones, I., Latinus, M., Fillion-Bilodeau, S., et al. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience*, *10*, 127.
- Cheour, M., Kushnerenko, E., Čeponienė, R., Fellman, V., & Näätänen, R. (2002). Electric brain responses obtained from newborn infants to changes in duration in complex harmonic tones. *Developmental Neuropsychology*, *22*, 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*, 4–16.
- Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., et al. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*, 53–58.
- Csibra, G., Kushnerenko, E., & Grossmann, T. (2008). Electrophysiological methods in studying infant cognitive development. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 247–262). Cambridge, MA: MIT Press.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*, 1174–1176.
- Decety, J. (2010). The neurodevelopment of empathy in humans. *Developmental Neuroscience*, *32*, 257–267.
- Decety, J. (2011). The neuroevolution of empathy. *Annals of the New York Academy of Sciences*, *1231*, 35–45.
- Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298*, 2013–2015.
- Dondi, M., Simion, F., & Caltran, G. (1999). Can newborns discriminate between their own cry and the cry of another newborn infant? *Developmental Psychology*, *35*, 418–426.
- Dunn, M. A., Gomes, H., & Gravel, J. (2008). Mismatch negativity in children with autism and typical development. *Journal of Autism and Developmental Disorders*, *38*, 52–71.
- Ecklund-Flores, L., & Turkewitz, G. (1996). Asymmetric headturning to speech and nonspeech in human newborns. *Developmental Psychobiology*, *29*, 205–217.
- Elsabbagh, M., & Johnson, M. H. (2007). Infancy and autism: Progress, prospects, and challenges. *Progress in Brain Research*, *164*, 355–383.
- Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., et al. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *NeuroReport*, *17*, 249–253.
- Ferri, R., Elia, M., Agarwal, N., Lanuzza, B., Musumeci, S. A., & Pennisi, G. (2003). The mismatch negativity and the P3a components of the auditory event-related potentials in autistic low-functioning subjects. *Clinical Neurophysiology*, *114*, 1671–1680.
- 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*, 238–252.
- Fox, N. A. (1991). If it's not left, it's right. Electroencephalograph asymmetry and the development of emotion. *American Psychologist*, *46*, 863–872.
- Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: From syllables to sentences. *Trends in Cognitive Sciences*, *9*, 481–488.
- Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *NeuroReport*, *13*, 1251–1254.
- Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., et al. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience*, *7*, 801–802.
- Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*, 145–146.
- Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, *28*, 219–236.

- Grossmann, T., & Friederici, A. D. (2011). When during development do our brains get tuned to the human voice? *Social Neuroscience*. doi: 10.1080/17470919.2011.628758.
- Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, *65*, 852–858.
- Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *NeuroReport*, *16*, 1825–1828.
- Herrmann, B., Maess, B., Hasting, A. S., & Friederici, A. D. (2009). Localization of the syntactic mismatch negativity in the temporal cortex: An MEG study. *Neuroimage*, *48*, 590–600.
- Hirasawa, K., Kurihara, M., & Konishi, Y. (2002). The relationship between mismatch negativity and arousal level. Can mismatch negativity be an index for evaluating the arousal level in infants? *Sleep Medicine*, *3*(Suppl. 2), S45–S48.
- Hobson, R. P., Ouston, J., & Lee, A. (1989). Naming emotion in faces and voices: Abilities and disabilities in autism and mental retardation. *British Journal of Developmental Psychology*, *7*, 14.
- Hutt, C., von Bernuth, H., Lenard, H. G., Hutt, S. J., & Prechtl, H. F. (1968). Habituation in relation to state in the human neonate. *Nature*, *220*, 618–620.
- Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *The Journal of Comparative Neurology*, *387*, 167–178.
- Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*, 220–224.
- Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage*, *22*, 948–955.
- Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*, R143–R145.
- Lloyd-Fox, S., Blasi, A., Mercure, E., Elwell, C. E., & Johnson, M. H. (2011). The emergence of cerebral specialization for the human voice over the first months of life. *Social Neuroscience*. doi: 10.1080/17470919.2011.614696.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Martin, G. B., & Clark, R. D. (1982). Distress crying in neonates: Species and peer specificity. *Developmental Psychology*, *18*, 3–9.
- Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology*, *35*, 204–214.
- 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*, 808–817.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoni, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178.
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-old infants prefer their native language. *Infant Behavior & Development*, *16*, 495–500.
- Nätänen, 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.
- Ockelford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of neonates to parents' and others' voices. *Early Human Development*, *18*, 27–36.
- Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience*, *11*, 367–374.
- Petkov, C. I., Logothetis, N. K., & Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? *Neuroscientist*, *15*, 419–429.
- Rogier, O., Roux, S., Belin, P., Bonnet-Brilhault, F., & Bruneau, N. (2010). An electrophysiological correlate of voice processing in 4- to 5-year-old children. *International Journal of Psychophysiology*, *75*, 44–47.
- Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, *32*, 189–194.
- Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., & Friederici, A. D. (2008). When vocal processing gets emotional: On the role of social orientation in relevance detection by the human amygdala. *Neuroimage*, *40*, 1402–1410.
- Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *NeuroReport*, *16*, 635–639.
- Van Lancker, D., Cornelius, C., & Kreiman, J. (1989). Recognition of emotional-prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuropsychology*, *5*, 20.
- Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*, 585–594.
- Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*, 437–456.