

Early Specialization for Voice and Emotion Processing in the Infant Brain

Anna Blasi,^{1,9,*} Evelyne Mercure,^{2,9,*} Sarah Lloyd-Fox,³ Alex Thomson,¹ Michael Brammer,⁴ Disa Sauter,⁵ Quinton Deeley,¹ Gareth J. Barker,⁴ Ville Renvall,⁶ Sean Deoni,^{4,7} David Gasston,⁴ Steven C.R. Williams,^{4,8} Mark H. Johnson,³ Andrew Simmons,^{4,8} and Declan G.M. Murphy^{1,8}

¹Department of Forensic and Neurodevelopmental Science, Institute of Psychiatry, King's College London, London SE5 8AF, UK

²Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, UK

³Centre for Brain and Cognitive Development, Birkbeck College, London WC1E 7HX, UK

⁴Department of Neuroimaging, Institute of Psychiatry, King's College London, King's Health Partners, De Crespigny Park, London SE5 8AF, UK

⁵Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands

⁶Brain Research Unit, Low Temperature Laboratory, Aalto University School of Science, FI-00076 Aalto, Espoo, Finland

⁷Advanced Baby Imaging Lab, School of Engineering, Brown University, Providence, RI 02912, USA

⁸NIHR Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, King's College London, London, SE5 8AZ, UK

Summary

Human voices play a fundamental role in social communication, and areas of the adult “social brain” show specialization for processing voices and their emotional content (superior temporal sulcus, inferior prefrontal cortex, premotor cortical regions, amygdala, and insula) [1–8]. However, it is unclear when this specialization develops. Functional magnetic resonance (fMRI) studies suggest that the infant temporal cortex does not differentiate speech from music or backward speech [9, 10], but a prior study with functional near-infrared spectroscopy revealed preferential activation for human voices in 7-month-olds, in a more posterior location of the temporal cortex than in adults [11]. However, the brain networks involved in processing nonspeech human vocalizations in early development are still unknown. To address this issue, in the present fMRI study, 3- to 7-month-olds were presented with adult nonspeech vocalizations (emotionally neutral, emotionally positive, and emotionally negative) and nonvocal environmental sounds. Infants displayed significant differential activation in the anterior portion of the temporal cortex, similarly to adults [1]. Moreover, sad vocalizations modulated the activity of brain regions involved in processing affective stimuli such as the orbitofrontal cortex [12] and insula [7, 8]. These results suggest remarkably early functional specialization for processing human voice and negative emotions.

Results

In this study, we used functional magnetic resonance imaging (fMRI) to investigate brain activation associated with the perception of adult nonspeech vocalizations by infants during natural sleep to address two specific aims: first, to determine whether the temporal cortex of young infants shows specialization for human voices, and second, to determine which brain areas are activated when human infants process emotion from nonspeech vocalizations. In adults, the human voice [1, 2] and the emotional information conveyed by human vocalizations have been found to modulate the activity of a number of brain areas, including the superior temporal sulcus (STS), inferior prefrontal cortex, premotor cortical regions, amygdala, and insula [4–8]. We investigated brain function in 21 infants aged 3–7 months who were presented with three types of adult nonspeech vocalizations (emotionally neutral, emotionally positive, and emotionally negative) and a mixture of nonvoice environmental sounds likely to be familiar to infants of that age. Patterns of fMRI activation representing the contrast between voice and nonvoice stimuli were studied, as well as the contrast between emotionally neutral, positive, and negative vocalizations.

Group analyses of all sounds versus rest (Figure 1; see also Table S1 available online) revealed significant activation in the middle temporal gyri, right lingual gyrus, medial frontal gyri, right putamen (lentiform nucleus), and right fusiform gyrus. This pattern of activation is consistent with reports of activation from auditory stimuli in other studies of infants, children, and adults [9, 10, 13, 14]. Hence, we can confirm that auditory activation was successfully recorded in these naturally sleeping infants.

Neutral vocalizations elicited more activation than nonvoice stimuli in the right anterior middle and superior temporal gyri (Figures 2A and 2B; Table S2) and in the medial frontal gyri (Figure 2A, in red). In contrast, the nonvoice stimuli elicited significantly more activation than neutral vocalizations in the left superior temporal gyrus (Figure 2A, in blue). A positive correlation in the voice versus nonvoice contrast was found between age and activation in a small cluster in the left superior temporal gyrus (cluster size = 9 voxels, correlation coefficient = 0.632, $p = 0.0006$; Figure 3; Table S3), suggesting that the voice selectivity increased with age in this area. No other area showed correlation with age.

No difference was observed between happy and neutral vocalizations. In contrast, sad vocalizations showed stronger activation than neutral vocalizations in the insula and gyrus rectus (Figure 4; Table S4).

Discussion

Voice-Sensitive Activation

Our results demonstrate a strong voice-sensitive area in the temporal cortex of 3- to 7-month-old infants. The strongest activation was found in the right middle temporal gyrus close to the temporal pole, in a location similar to the anterior portion of the voice-sensitive area reported in adults [1, 2, 15–18]. This area along the anterior superior temporal sulcus (STS) shows

⁹These authors contributed equally to this work

*Correspondence: anna.blasi@kcl.ac.uk (A.B.), e.mercure@ucl.ac.uk (E.M.)

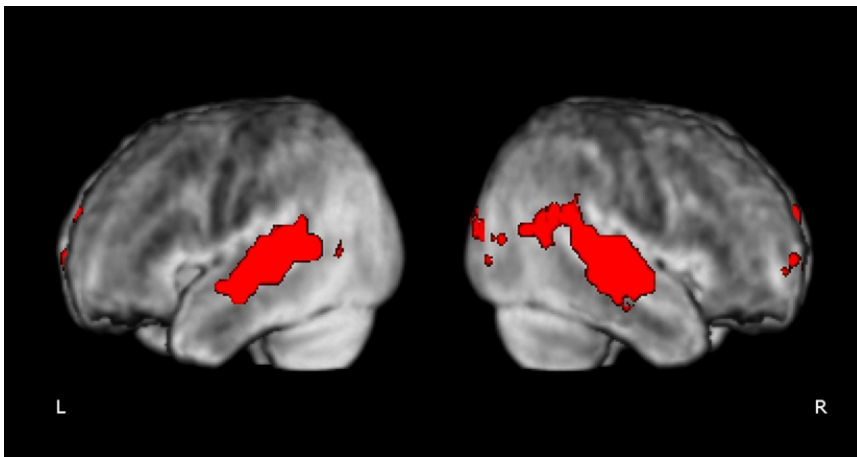


Figure 1. All Sounds versus Rest Condition

Three-dimensional representation of activation to all sounds compared with rest. The activations are significant at $p \leq 0.005$. L = left; R = right. See also Table S1.

more activation in response to human vocalizations (speech and nonspeech) than other sounds, including nonvocal environmental sounds [1] and animal vocalizations [19]. It also shows adaptation to the speaker's identity in adults [15] and activates more strongly when participants focus on the speaker's voice rather than on the verbal content of spoken sentences [20], regardless of whether these voices are familiar or not [21]. Because the stimuli were the same and only the

focus of the participant's attention differed, this result suggests that the activation of this STS area cannot be attributed to low-level differences in stimulus categories. For these reasons, this area of the anterior STS has been described as having an important role in the processing of human voices, especially the identification of speakers [15]. In contrast, Grossmann et al. [11] found voice sensitivity in a more posterior temporal area in 7-month-olds, which

has been interpreted as an indication of a developmental difference in the voice sensitivity of more anterior temporal areas [20]. The current results suggest that this localization discrepancy between prior studies of infants and adults is more likely to reflect differences in brain imaging techniques (for example, spatial resolution), than genuine developmental differences.

In our study, the voice-sensitive activation was right lateralized in the temporal cortex, which is congruent with previous

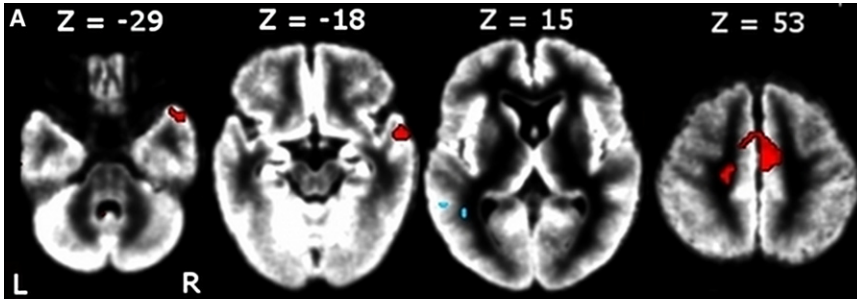
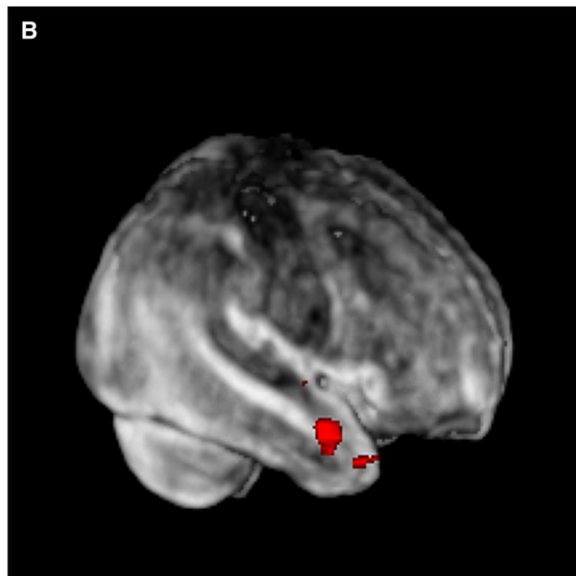


Figure 2. Neutral versus Nonvoice Contrast

Representation on a gray-matter infant template of the results for the contrast between neutral and nonvoice conditions.

(A) Neutral vocalization > nonvoice (in red), neutral vocalization < nonvoice (in blue). The Talairach z coordinates mark the position of the voxel with maximum activation.

(B) Three-dimensional rendering of the contrast neutral vocalization > nonvoice, showing the right anterior temporal activation. See also Table S2.



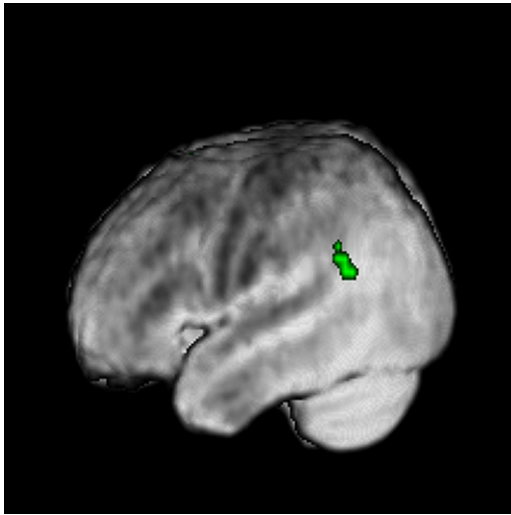


Figure 3. Correlation with Age
Representation of the cluster in the left superior temporal gyrus (Brodmann area 22) showing positive correlation with age for the contrast neutral vocalization versus nonvoice. See also Table S3.

functional near-infrared spectroscopy (fNIRS) findings in babies [11] and fMRI findings in adults [19, 22]. The right anterior superior temporal sulcus has been specifically related to the analysis of nonverbal features of speech, showing more activation when focusing on the speaker's voice than when focusing on the semantic content of spoken sentences [20]. Our results suggest that a right-hemisphere bias for processing human vocalizations emerges early in development.

Although no correlation was found with age for the main effects in the human voice versus environmental sounds contrast, a small volume in the left superior temporal gyrus did show a positive correlation with age. This small volume was localized in a similar region to findings from an infant fNIRS study that reported age-correlated differences in activation for human voice [11]. Interestingly, whereas the current findings were left lateralized, the findings from the fNIRS

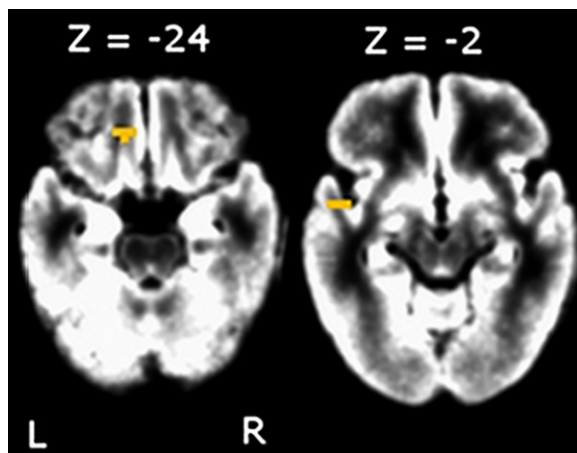


Figure 4. Sad versus Neutral Vocalizations
The clusters represent increased activation with sad compared with neutral vocalizations. The corresponding Talairach z coordinate is given above each slice. See also Table S3.

studies suggested bilateral age-correlated changes in activation in the temporal cortex. The sleeping state of the infants could have influenced our findings by reducing the blood oxygen level-dependent contrast responses as observed in adults [23] or by reducing the differentiation between responses to different stimuli as observed in babies [9]. Also, in fNIRS studies with infants, it is necessary to use multimodal stimuli (visual and auditory) to maintain the baby's interest in the task [24]. This multimodal stimulus presentation may also have influenced the responses recorded with fNIRS, further accentuating the differences with our findings.

Emotion-Modulated Activation

Our second aim was to explore brain areas in which activation is modulated by the emotional content of human vocalizations in young infants. Many studies with adults have found that the emotional information conveyed by human vocalizations modulates the activity of certain brain areas, including the temporal voice-sensitive area, inferior prefrontal cortex, premotor cortical regions, amygdala, and insula [4–8]. A prior study of infants using fNIRS [11] also found that emotional prosody in spoken words increased the activation of the voice-sensitive area of the right temporal hemisphere in 7-month-olds. In the same age group, event-related potentials (ERPs) revealed a positive slow wave over temporal electrodes when words were spoken with a happy or angry prosody but not with a neutral prosody [25]. However, ERPs have a very low spatial resolution, and one technical limitation of fNIRS is that it can only be used to measure the activity in superficial areas of the cortex. In addition, the technique is limited by the placement of the fNIRS probes, which in the aforementioned study was over the inferior frontal and temporal cortex [11]. Therefore, a unique contribution of our study is the investigation of emotion within voice over the whole infant brain using an imaging technique of high spatial resolution. Moreover, in contrast to studies on speech, the present study used emotional signals that preverbal infants produce themselves from a very early age (crying and laughter) and that are thought to reflect innate behaviors to communicate emotional states [26].

In the present study, the activation elicited by happy vocalizations did not differ from that of emotionally neutral vocalizations, whereas sad vocalizations elicited significantly greater responses than emotionally neutral vocalizations in the insula and orbitofrontal cortex. These results are congruent with findings of increased activation in the insula when adults listen to emotionally salient nonspeech vocalizations (especially sad and fearful vocalizations) [7, 8]. The orbitofrontal cortex has been described as one of the “least understood regions of the human brain” [12]. Available data suggest a role of this area in the processing of affective stimuli [12], such as recognizing emotions from facial expressions [27], although little is known about the maturation processes of the orbitofrontal cortex in childhood and adolescence [12]. The present study suggests that the role of the orbitofrontal cortex for processing emotions extends to emotions presented in the auditory modality and emerges early in human development.

The temporal cortex area identified as voice selective in the present study did not show any modulation of activity based on the emotional content of the stimuli. This is in contrast with other studies using fNIRS [11] that found an increase of activity with emotional prosody in infants in the voice-sensitive temporal channels, and others using fMRI [3] that found enhanced response of the temporal voice-sensitive area with

words spoken with an angry prosody in adults. The relative lack of differences between the emotional (especially happy) and neutral conditions may reflect the fact that the participants in the present study were asleep, therefore these responses may have been more difficult to differentiate [9], whereas they were awake in previous studies of emotional vocalizations. Moreover, the “neutral” vocal stimuli in the present study (coughing, sneezing, throat clearing) may be emotionally salient to infants because they indicate the presence of adults. In adult studies, classification of stimuli as neutral or emotional is based on participants’ rating (e.g., [28]), but an infant’s perception of their emotional content may differ. The terms “sad,” “happy,” and “neutral” used here should be taken as labels to describe the stimuli as perceived by adults and do not imply that the same emotions are evoked in infants when hearing these stimuli. Furthermore, the finding that activation was evident for sad versus neutral but not happy versus neutral vocalizations could be an effect of unfamiliarity, because young infants are likely to be less frequently exposed to crying than adult laughter and neutral vocalizations. The same activation may not be observed for sad vocalizations when studying the response to the sound of a child crying in a group of infants who have older siblings or who attend childcare settings from an early age. Studying the infants of depressed mothers may also be a way of assessing the role of early experience, because these infants may be exposed to an atypical balance of neutral, happy, sad, and angry vocalizations. Nevertheless, our findings indicate that emotional vocalizations are associated with differential activity in emotion-processing networks from an early age.

In conclusion, voice-sensitive activation was observed along the anterior STS in 3- to 7-month-olds in a location very similar to that described in the adult brain. This suggests that the infant temporal cortex shows more refined functional specialization than previously reported. This early functional specialization for processing the human voice parallels infants’ surprisingly early ability to extract subtle information from human vocalizations. Indeed, newborns prefer to listen to their mother’s voice [29] and their mother tongue [30], and young infants can also discriminate emotional vocalizations [31] and can differentiate male versus female and child versus adult vocalizations [32]. Moreover, the emotional content of vocalizations, especially sadness, may modulate the activity of areas known to be involved in processing affective stimuli in the adult brain, such as the orbitofrontal cortex and insula. Unresolved questions for the future include how this functional specialization is influenced by prenatal and early postnatal experience and whether these processes are disrupted in developmental disorders such as autism. Nevertheless, these results represent a rare demonstration that cortical specialization exists very early in development, a fundamental advance in our understanding of infant development. The findings of this study also highlight the potential of fMRI as a tool for studying the development of specific brain responses to auditory stimuli in infancy.

Experimental Procedures

Participants

Data from 21 naturally sleeping infants were included in the analysis (age range 91–203 days, mean age 152 ± 30 days, gestation-corrected). All parents gave written informed consent. The study was approved by the Institute of Psychiatry and South London and Maudsley research ethics committee.

Stimuli

Infants were presented with four categories of auditory stimuli: neutral vocalizations (natural nonspeech vocalizations with no strong emotional content, such as coughing, sneezing, yawning, and throat clearing), happy vocalizations (laughing), sad vocalizations (crying), and nonvoice sounds (environmental sounds that are not human or animal produced but are likely to be familiar to infants of that age, such as toy sounds and water sounds). All voice sounds were adult vocalizations. Each condition was presented for 21 s, with an interval of 9 s of rest, organized in a block design.

Data Analysis

The MRI data were analyzed with XBAM software (www.brainmap.co.uk/xbam.htm) using a data-driven approach based on the standard general linear model. Data were normalized to Talairach space using an infant template previously described by Dehaene-Lambertz et al. [9]. Statistical analysis was performed using a random-effects model, and only clusters of at least three voxels in size are reported.

Supplemental Information

Supplemental Information includes two figures, four tables, and Supplemental Experimental Procedures and can be found with this article online at [doi:10.1016/j.cub.2011.06.009](https://doi.org/10.1016/j.cub.2011.06.009).

Acknowledgments

Funding was provided by the Medical Research Council (MRC UK), the MRC UK AIMS network (G0400061/69344, D.G.M.M., principal investigator), and the NIHR Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, King’s College London. M.H.J. and S.L.-F. were supported by MRC grant G0701484. We would like to thank Ghislaine Dehaene-Lambertz for her valuable input on infant MRI scanning and the team of Joe Piven.

Received: February 11, 2011

Revised: May 9, 2011

Accepted: June 6, 2011

Published online: June 30, 2011

References

1. Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312.
2. Belin, P., Zatorre, R.J., and Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Res. Cogn. Brain Res.* 13, 17–26.
3. Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., and Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nat. Neurosci.* 8, 145–146.
4. Fecteau, S., Armony, J.L., Joanette, Y., and Belin, P. (2005). Sensitivity to voice in human prefrontal cortex. *J. Neurophysiol.* 94, 2251–2254.
5. Warren, J.E., Sauter, D.A., Eisner, F., Wiland, J., Dresner, M.A., Wise, R.J.S., Rosen, S., and Scott, S.K. (2006). Positive emotions preferentially engage an auditory-motor “mirror” system. *J. Neurosci.* 26, 13067–13075.
6. Fecteau, S., Belin, P., Joanette, Y., and Armony, J.L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage* 36, 480–487.
7. Morris, J.S., Scott, S.K., and Dolan, R.J. (1999). Saying it with feeling: Neural responses to emotional vocalizations. *Neuropsychologia* 37, 1155–1163.
8. Sander, K., and Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Brain Res. Cogn. Brain Res.* 12, 181–198.
9. Dehaene-Lambertz, G., Dehaene, S., and Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015.
10. Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allior, L., Dubois, J., Hertz-Pannier, L., and Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants’ language networks. *Brain Lang.* 114, 53–65.
11. Grossmann, T., Oberecker, R., Koch, S.P., and Friederici, A.D. (2010). The developmental origins of voice processing in the human brain. *Neuron* 65, 852–858.

12. Kringelbach, M.L. (2005). The human orbitofrontal cortex: Linking reward to hedonic experience. *Nat. Rev. Neurosci.* **6**, 691–702.
13. Altman, N.R., and Bernal, B. (2001). Brain activation in sedated children: Auditory and visual functional MR imaging. *Radiology* **221**, 56–63.
14. Redcay, E., Kennedy, D.P., and Courchesne, E. (2007). fMRI during natural sleep as a method to study brain function during early childhood. *Neuroimage* **38**, 696–707.
15. Belin, P., and Zatorre, R.J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* **14**, 2105–2109.
16. Belin, P., Fecteau, S., and Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends Cogn. Sci. (Regul. Ed.)* **8**, 129–135.
17. Latinus, M., and Belin, P. (2011). Human voice perception. *Curr. Biol.* **21**, R143–R145.
18. Scott, S.K. (2008). Voice processing in monkey and human brains. *Trends Cogn. Sci. (Regul. Ed.)* **12**, 323–325.
19. Fecteau, S., Armony, J.L., Joanette, Y., and Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage* **23**, 840–848.
20. von Kriegstein, K., Eger, E., Kleinschmidt, A., and Giraud, A.L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res. Cogn. Brain Res.* **17**, 48–55.
21. Kriegstein, K.V., and Giraud, A.L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* **22**, 948–955.
22. Belin, P., and Grosbras, M.-H. (2010). Before speech: Cerebral voice processing in infants. *Neuron* **65**, 733–735.
23. Czisch, M., Wetter, T.C., Kaufmann, C., Pollmächer, T., Holsboer, F., and Auer, D.P. (2002). Altered processing of acoustic stimuli during sleep: Reduced auditory activation and visual deactivation detected by a combined fMRI/EEG study. *Neuroimage* **16**, 251–258.
24. Lloyd-Fox, S., Blasi, A., and Elwell, C.E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neurosci. Biobehav. Rev.* **34**, 269–284.
25. Grossmann, T., Striano, T., and Friederici, A.D. (2005). Infants' electric brain responses to emotional prosody. *Neuroreport* **16**, 1825–1828.
26. Barr, R.G., Hopkins, B., and Green, J.A. (2000). *Crying as a Sign, a Symptom, & a Signal* (New York: Cambridge University Press).
27. Leppänen, J.M., and Nelson, C.A. (2009). Tuning the developing brain to social signals of emotions. *Nat. Rev. Neurosci.* **10**, 37–47.
28. Belin, P., Fillion-Bilodeau, S., and Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* **40**, 531–539.
29. DeCasper, A.J., and Fifer, W.P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science* **208**, 1174–1176.
30. Moon, C., Cooper, R.P., and Fifer, W.P. (1993). Two-day-olds prefer their native language. *Infant Behav. Dev.* **16**, 495–500.
31. Walker-Andrews, A.S., and Lennon, E.M. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behav. Dev.* **14**, 131–142.
32. Bahrick, L.E., Netto, D., and Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Dev.* **69**, 1263–1275.