

Language or Music, Mother or Mozart?

Structural and environmental influences on infants' language networks

G. Dehaene-Lambertz¹⁻⁴, A. Montavont¹, A. Jobert^{1,3}, L. Alliol³⁻⁴, J. Dubois³⁻⁴, L. Hertz-Pannier³⁻⁵, S. Dehaene^{1,3}

¹INSERM, U562, Neurospin, 91191 Gif/Yvette, France, Université Paris 11, Paris, France

²AP-HP, service de Neurologie Pédiatrique, CHU Bicêtre, Le Kremlin-Bicêtre, France

³IFR49, Neurospin, 91191 Gif/Yvette, France

⁴CEA, Laboratoire de recherche biomédicale, CEA/SAC/DSV/I2BM/NeuroSpin, Gif-sur-Yvette, France

⁵INSERM, U663, Paris, France; Université Paris5, Paris, France

Corresponding author: G Dehaene, Unité INSERM 562, CEA/SAC/DSV/DRM/NeuroSpin, Bat145, point courrier 156, 91191 GIF/YVETTE, France. E-mail: ghislaine.dehaene@cea.fr, Telephone: 33 1 69 08 81 72; Fax: 33 1 69 08 79 73

Abstract

Understanding how language emerged in our species calls for a detailed investigation of the initial specialization of the human brain for speech processing. Our earlier research demonstrated that an adult-like left-lateralized network of perisylvian areas is already active when infants listen to sentences in their native language, but did not address the issue of the specialization of this network for speech processing. Here we used fMRI to study the organization of brain activity in 2-month-old infants when listening to speech or to music. We also explored how infants react to their mother's voice relative to an unknown voice. The results indicate that the well-known structural asymmetry already present in the infants' posterior temporal areas has a functional counterpart: there is a left-hemisphere advantage for speech relative to music at the level of the *planum temporale*. The posterior temporal regions are thus differently sensitive to the auditory environment very early on, channelling speech inputs preferentially to the left side. Furthermore, when listening to the mother's voice, activation was modulated in several areas, including areas involved in emotional processing (amygdala, orbitofrontal cortex), but also, crucially, a large extent of the left posterior temporal lobe, suggesting that the mother's voice plays a special role in the early shaping of posterior language areas. Both results underscore the joint contributions of genetic constraints and environmental inputs in the fast emergence of an efficient cortical network for language processing in humans.

Key-words: fMRI, brain, infant, language acquisition, lateralization, social, music

Introduction

Infants present early capacities to process speech and to rapidly learn properties of their native language (Dehaene-Lambertz, Hertz-Pannier, & Dubois, 2006; Jusczyk, 1997; P. K. Kuhl, 2004; Werker & Curtin, 2005). Thanks to the development of non-invasive brain imaging techniques that can be safely used with human infants, the cerebral bases underlying these early stages of language acquisition can now be investigated. It becomes possible to ask, in a purely empirical manner, a question which was previously only examined through indirect theory-driven speculation: what is the particular organization of the human brain that permits language learning in our species?

In previous papers, using fMRI, we discovered that when infants listen to speech in their native language, activation was not distributed widely to a broad set of areas, but was already concentrated to a set of left-hemispheric perisylvian regions similar to that found in human adults (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002). Furthermore we demonstrated that this network is already functionally organized, as the different brain regions involved are sensitive to different properties of the stimuli or of the tested paradigm. For example, some regions, including the angular gyrus and the precuneus, discriminate forward from backward speech, while temporal regions do not (Dehaene-Lambertz et al., 2002). Furthermore, the left inferior frontal region is sensitive to sentence repetition, suggesting its participation in an early verbal working memory system (Dehaene-Lambertz, Dehaene et al., 2006). We also observed a temporal gradient in the activation produced by the perception of a single sentence: the BOLD response is increasingly delayed as one moves from the auditory primary cortex to the posterior part of the superior temporal gyrus and to the temporal poles and inferior frontal regions (Broca's area). The delays, again paralleling those found in the adult speech network, suggest a hierarchical organization of the human temporal lobes which presents homologies with that of the monkey brain. We speculated that it might be useful to learn about the nested structure of speech, which contains temporal regularities at multiple hierarchically organized time scales (Dehaene-Lambertz, Hertz-Pannier, Dubois, & Dehaene, 2008).

These studies, however, did not reveal which aspects of this organization, if any, are specific for speech, and which would be equally engaged by equally structured and complex acoustic stimuli outside of the speech domain. One particular question concerns the asymmetry of the linguistic network and whether there exists an early left-hemispheric bias for processing speech stimuli. Several findings point to large differences in the maturation of

the left and right hemispheres in humans. First, genetic studies have revealed asymmetrical gene expression in the perisylvian regions, specific to the human lineage (Sun, Collura, Ruvolo, & Walsh, 2006). This is observed especially between 12 and 14 weeks of gestation, a critical time for cortical regionalization. Second, gyration develops earlier over the right hemisphere than over the left (Chi, Dooling, & Gilles, 1977; Dubois, Benders et al., 2008). Third, the significant anatomical asymmetries which are present at the level of the planum temporale and of the superior temporal sulcus in human adults (Van Essen, 2005) are already clearly observed during the first months of life (Glasel, 2007). Finally, analyses of a fiber bundle of major linguistic relevance, the arcuate fasciculus that links temporal and inferior frontal regions, reveal that left-right differences are already present in 1 to 4 month-old infants (Dubois, Hertz-Pannier et al., 2008). At this age, the left temporal part of this bundle is already larger than the right, and maturation is more advanced in its left parietal part than on its right, as assessed by fractional anisotropy in diffusion tensor imaging.

From a functional point of view, the few existing brain imaging studies in the first months of life also suggest an early functional asymmetry. The amplitude of event-related potentials to auditory stimuli is larger over the left hemisphere than the right in two-month-old infants (Dehaene-Lambertz, 2000), and the hemodynamic response to sentences in the native language is significantly asymmetric in the posterior superior temporal region in neonates (Pena et al., 2003) and three-month-olds (Dehaene-Lambertz et al., 2002; Dehaene-Lambertz, Hertz-Pannier, Dubois et al., 2006). However, from these studies alone, it is not possible to determine if these functional asymmetries reflect a domain-general advantage of the left hemisphere for auditory processing, or a domain-specific advantage of the left perisylvian regions for speech processing. Very few experiments point to a genuine functional differentiation of the left and right hemispheres with respect to language processing. Using high-amplitude sucking in a dichotic listening paradigm, Bertoncini et al (1989) observed that neonates react more to a change of syllable in the right than in the left ear while the reverse preference was observed for a change of timbre. These findings supported Entus's (1977) early results obtained in 1-4 month-old infants. However, lateralized preferences for linguistic and non linguistic stimuli were not so clear-cut in several other studies performed during the first months of life (Best, Hoffman, et Glanville, 1982; Glanville, Best, et Levenson, 1977; Vargha-Khadem et Corballis, 1979). In sleeping 3-month-old infants, Homae et al (2006) recorded with near infra-red spectroscopy (NIRS) a larger right temporo-parietal response for normal sentences compared to sentences without prosody while the responses were similar for both types of sentences over the left hemisphere. However, the authors did not mention a

significant interaction between hemisphere and type of stimuli. Bristow et al (Bristow et al., in press) used high-density ERPs to study 3-month-old infants' responses to crossmodal auditory and visual representations of gender and vowel identity. Infants were presented with male or female faces that produced a silent articulatory gesture, followed by an auditory vowel. The vowel was either congruent with the previous articulatory movement or not, and it was produced by a speaker whose gender was congruent or not with the gender of the previous faces. Incongruence for gender and for vowel identity induced very different mismatch responses on the scalp. For a vowel change, the sources of the mismatch responses were clearly left lateralized, while it was right lateralized for a gender change. This asymmetry suggests an early functional segregation of processing between right and left hemisphere, similar to what is described in adults.

These previous studies, however, used techniques of dichotic listening or event-related potentials that only provided an indirect reflection of the underlying functional localization. If NIRS can provide better spatial resolution, the exact localization of the cerebral response is not known. In the present study, we sought to examine the functional specialization of infants' cortical networks using the anatomically accurate method of fMRI. To test whether a left advantage is observed only for speech or for any organized stimulus, we presented speech and music segments to two-month-old infants. Furthermore, to probe the possible impact of environmental learning on this initial organization, within the speech condition we contrasted the mother's voice with an unknown female voice (the mother of the preceding baby in the study). In adults, the left-hemispheric lateralization of linguistic processing has been classically opposed to a right-hemispheric advantage for voice identification and discrimination, and for emotional content deciphering (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Bowers, Coslett, Bauer, Speedie, & Heilman, 1987; Ethofer et al., 2006). The ERP experiment cited above on cross-modal representation of vowel and gender (Bristow et al., in press) suggests that a similar dissociation may exist in infants. Using two different voices, one of which was highly known and emotionally salient, we aimed to explore the activations induced by the mother's voice above and beyond those induced by speech processing alone, possibly uncovering right activations for some aspects of its perception.

Finally, our experimental design included a supplementary factor of sentence repetition. Segments were either constituted of 12 seconds of speech/music sentences, or of a single four-second segment that was repeated three times. In human adults, repetition of the

same stimulus classically induces a reduction in activation (repetition suppression), whereas repetition enhancement can sometimes be observed when the subject actively tracks the repeated stimulus. When such repetition effects are present in a brain region, they indicate that this region contains a representation of the past stimulus, and variants of the repetition paradigm can then be used to monitor the abstractness of this representation (Grill-Spector & Malach, 2001; Naccache & Dehaene, 2001). In a previous infant fMRI study, using a slow event-related paradigm where a single sentence was repeated at a 14.4 s interval, we did not observe any repetition suppression, but repetition enhancement was observed in the left inferior frontal region. The absence of repetition suppression in the temporal region was surprising, because it is a reliable effect observed in adults when the same sentence is repeated (Dehaene-Lambertz, Dehaene et al., 2006), and even in infants at birth, repetition of a syllable every 600 ms produces an effective decrease in ERPs amplitude (Dehaene-Lambertz & Dehaene, 1994; Dehaene-Lambertz & Pena, 2001). This absence could have been related to the particularities of the BOLD response in infants, to the complexity of having to remember a full sentence, or to the large time-lag between successive sentences, possibly erasing the echoic buffer of the temporal regions. An immediate repetition as in the present experiment should help to disentangle these different hypotheses and to clarify the characteristics of auditory representations in the temporal lobe.

Method

Subjects

Healthy full-term infants were tested around two and a half months of age. All parents gave their written informed consent for the protocol. No sedation was used. The study was approved by the local ethical committee for biomedical research (CCPPRB, Hôpital de Bicêtre, Le Kremlin-Bicêtre, France).

Among the 24 two-month-old infants who came to the lab, we obtained useful functional data in only 7 infants (5 males and 2 females, mean age 72 days, 59 to 88 days). In 6 infants the exam was not started or was stopped before the acquisition of any functional data because of fussiness. In the other 11 babies, no significant activation was observed for the overall sound versus silence contrast, even at $p=0.05$ in auditory regions. These experimental failures might be imputed to a variety of factors: small number of acquired images in these

infants; large amplitude of their movements when they were awake; reduced or absent cortical responsiveness in deeply asleep infants (3 infants); or a reduced or absent BOLD response, which is thought to quickly change and even reverse sign at around this age (Yamada et al., 1997). This high attrition rate underscores the fact that fMRI remains a challenge at this age. In sedated children tested between 2 months and 9 years, Altman and Bernal (2001) reported auditory activation in only 68% of the 38 patients listening to their mother's voice. In the younger group (< 1 year), almost half of the infants had no auditory responses (5/11) in the temporal regions, although occipital responses to flashing lights were present in 6 out of the 7 infants tested. Auditory stimuli therefore seem to pose a special challenge, perhaps because they are not salient enough to elicit activity in deeply asleep infants.

Stimuli

Three types of stimuli (music, mother's and stranger's speech) were recorded in two different forms (repeated or varied conditions). Five different occurrences of about 12 seconds were created for each condition. In the repeated conditions, a segment of approximately four seconds was repeated three times while in the non-repeated conditions, an entire segment of approximately twelve seconds was constituted of different sentences. Thus, the total duration was matched across conditions.

Krumhansl and Jusczyk (1990) have shown that infants are sensitive to the phrase structure of Mozart's piano pieces, just like they are sensitive to sentence structure in speech (Hirsh-Pasek et al., 1987). Guided by these earlier behavioural results, for the music condition we choose ten passages extracted from Mozart's piano sonata and containing entire musical phrases with natural beginning and ending (five 12-second-long and five 4-second-long passages). For the speech condition, prior to the neuroimaging session, the mother was recorded reading a child book. The same book, unknown to the mothers and thus to the infants, was used for all participants. The mother was told to read it as if she was reading to her infant. From her record, five 12-second and five 4-second passages were excerpted. The passages contained only full sentences. For each infant, the recordings from the previous mother were considered as the stranger condition. Thus, over the entire group of infants, the very same stimuli were used in the Mother and Stranger conditions.

Procedure

Scanning was performed on a 1.5 Tesla MR scanner (GE medical system). While the infant was quiet, noise protection earphones with inserted piezoelectric loudspeakers were placed, with sides reversed for each successive infant, then the infant was placed in the scanner. To maintain quietness, a mirror placed above the infants' head allowed them to watch the outside of the tunnel. Toys were presented to keep them quiet by an experimenter who stayed inside the magnet room and was unaware of the stimuli presented to the infant. The experimenter checked the infant's wakefulness and comfort throughout the study and stopped the experiment at any point if the infant showed discomfort. Vigilance was ascertained on the basis of visual inspection.

1) Stimulation procedure and imaging parameters

We used a block-design with 12 seconds of sound followed by 7.2 seconds of silence. The six conditions were randomly presented one time during each fMRI run and infants got in average 10.14 runs (9 to 11). The level of sound presentation was adjusted to a comfortable level, easily understandable above the residual scanning noise by a normal adult. Only one infant stayed awake during the entire experiment while two infants were already asleep at the beginning of the functional sequences. The four others fell asleep at different moment during the experiment, respectively after 3, 1, 5 and 3 functional blocks. From visual inspection, infants were judged undoubtedly asleep during 48 runs over the 71 acquired runs (67.7% of the total number of runs), but it is difficult to judge how drowsiness might have affected infant's attention toward the stimuli during the preceding blocks. Because of the small number of subjects, it was not possible to separate awake and asleep infants in this study, and the data was therefore pooled.

T2-weighted spin-echo images were acquired for anatomical reference (24 contiguous axial slices of 4 mm thickness, 256 x 192 matrix, voxel size 0.977 x 0.977 x 4 mm³, TR=4000 ms, TE=120 ms) at the beginning of the MRI testing. The selected slices covered 9.6 cm and systematically included the cerebellum and occipital regions. The same volume was then imaged with a gradient-echo EPI sequence (16 axial slices of 5.5 mm thickness with a gap of 0.5 mm, 64 x 64 matrix, voxel size = 3.75 x 3.75 x 5.5 mm³, TR = 2400 ms, TE = 60 ms). 52 EPI volumes were acquired on each sequence. Four initial dummy scans, corresponding to an initial silent period of 9.6 s, were used to achieve steady-state magnetization.

2) Data processing and analysis

Data processing was performed using a customized version of Statistical Parametric Mapping software (SPM99). Each EPI volume was visually examined and volumes with severe movement artefacts were rejected. On average 477 volumes per infant were kept for analysis (range 432-528). For movement correction, all EPI volumes were realigned to an EPI volume devoid of artefacts, using a six-parameter linear transform. EPI volumes were then coregistered with the subject's anatomical T2 image using SPM's default coregistration algorithm. The quality of the coregistration was visually inspected and manually corrected if necessary. Data were then normalized to the infant template described in (Dehaene-Lambertz et al., 2002) by first computing a non-linear normalization transformation on the infant's anatomical image, then applying the same transformation to each EPI volume. For localization, activations were superimposed on a $1 \times 1 \times 1 \text{ mm}^3$ T2-weighted anatomical image acquired in an infant and registered to the infant template. EPI volumes were resampled using a sinc function and $4 \times 4 \times 4 \text{ mm}^3$ voxels, then smoothed using a 5 mm Gaussian kernel.

We generated a general linear model with six conditions (3 types of stimuli * repeated or not), convolved by the standard adult hemodynamic response function (HRF). To accommodate the high level of motion in infant EPI data, the six movement parameters (3 rotations, 3 translations) were entered as regressors of non-interest capable of capturing motion-induced changes in BOLD signal intensity. Second, SPM software was modified to allow voxels to stay in the analysis even if there was an occasional motion-induced signal loss (on at most 10 % of volumes). Third, deviations of the signal of more than 2.5 standard deviations (measured within each voxel's time series) were clamped so that they did not exceedingly disrupt the fit of the general linear model. Fourth, the linear model of the BOLD response was adapted for temporal sequences with occasional missing data.

Random-effect analyses:

At the group level, to partially compensate for anatomical variability and improve inter-subject activation overlap, individual contrasts from the HRF model were first resmoothed with a 5 mm gaussian kernel, then submitted to one-sample statistics with subjects as the random variable. Activations are reported when voxels were significant at $p < 0.01$ and formed a contiguous cluster whose extent was significant at $p < 0.05$, corrected for multiple comparisons across the brain volume.

To study whether functional asymmetries were different for speech and music, the smoothed individual images for the three contrasts (mother's voice, stranger's voice and

music) were flipped by applying a spatial transformation matrix, based on SPM's spatial realignment procedure, which used linear and non-linear terms to optimally align the left hemisphere of the infant's template onto the right and vice-versa (Dehaene-Lambertz et al., 2002). Individual asymmetry images were obtained by subtracting the original contrast image from its flipped version. Those images were then entered into a paired t-test analysis which tested whether the amount of activation for the studied contrast was significantly larger in one hemisphere relative to the other (voxel $p < 0.01$, cluster $p < 0.05$ corrected).

Analyses of regions of interest:

Random-effect analyses were completed by ROI analyses to study the differences between the three types of stimuli within the clusters identified by the previous whole-brain analyses. We used the MarsBar software package to extract the activation values averaged across the voxels of the identified clusters for each individual contrast image. Mean activation was then entered into an ANOVA with subjects as the random variable and type of stimuli and hemisphere as within-subjects variables. We only used this approach to evaluate contrasts orthogonal to, and thus statistically independent from, those originally used to define these clusters.

We also defined an a-priori region, the *planum temporale*, as a specific target structure to investigate the lateralization of speech and music. The *planum temporale* is classically reported as an asymmetric structure both at an anatomical level and at a functional level in adults (Tzourio, Nkanga-Ngila, & Mazoyer, 1998) and we observed significant functional asymmetries around the posterior temporal region in our two previous fMRI studies (Dehaene-Lambertz et al., 2002; Dehaene-Lambertz, Hertz-Pannier, Dubois et al., 2006). However, because of the shorter and steeper sylvian scissure on the right side already present in infants (Sowell et al., 2002), it cannot be excluded that the observed functional asymmetries may have been related to a systematic local shift imperfectly corrected by brain normalization and flipping of the left-right axis. Therefore, we took advantage of our ongoing anatomical research on the infant brain to create precisely defined anatomical regions of interest independently drawn in the left and right hemisphere. In another group of 14 infants (3 to 16 weeks-old), both *planum temporale* were manually drawn on high resolution T2 images (0.8 x 0.8 x 0.8 mm) from the posterior border of Heschl's gyrus up to the highest point of the most posterior branch of the sylvian scissure (Glaser et al., in preparation). Individual 3D meshes of the structure were normalized using the transformations created for the corresponding individual anatomy toward our infant's template, then smoothed with a 5

mm Gaussian kernel and finally averaged across subjects in order to create a probability map. To avoid discontinuities, this map was smoothed with a 8 mm Gaussian kernel and used to create a binary image of the left and right *planum temporale* (left = 4094 mm³ and right = 4847 mm³). We then extracted the activation values averaged across the voxels present in this image for each individual contrast image using the MarsBar software package and entered them in an ANOVA with within-subject factors of Stimulus Type (Mother's voice, Stranger's voice and Music) and Hemisphere (left and right).

Fixed-effect analyses:

Random-effect analyses did not reveal any significant differences between mother's and stranger's voice conditions, yet infants of this age are known to be able to discriminate these two conditions (DeCasper & Fifer, 1980; Mehler, Bertoncini, & Barriere, 1978). Random-effect analyses may have failed because the small number of subjects in our study required voxels to exhibit a similar response across essentially all subjects and with a low variance. We reasoned that any heterogeneity of responses across subjects, possibly due to the difficulty of maintaining a fixed level of vigilance, attention and cooperation (e.g. almost all the infants fell asleep), may have masked a subtle brain activation difference between these two intonated female voices. Therefore, to determine whether there was indeed no difference between the two voice conditions, we also performed multi-subjects fixed-effect analyses, which are more resistant to an imperfect spatial correspondence between subjects and to a poor signal-to-noise ratio that might exist in some subjects. For these analyses, activations are reported when voxels were significant at $p < 0.001$ and formed a contiguous cluster whose extent was significant at $p < 0.05$, corrected for multiple comparisons across the brain volume.

Results

Random-effect analyses :

The analysis of the activations induced by sound perception revealed significant bilateral activations in the left and right superior temporal sulci and gyri (table 1).

Repetition effect

A repetition suppression effect was observed in the left superior temporal gyrus extending toward the superior temporal sulcus and the middle temporal gyrus (table 1 and figure 1). The random analysis of the difference between activation in the left and right hemispheres did not identify any significant asymmetry for the repetition contrast.

ROI analysis: Activation averaged across the voxels of this left cluster and across the voxels of the contralateral region was extracted from each individual contrast image and entered in an ANOVA with type of stimuli (3 levels), repetition (2 levels) and hemisphere (2 levels) as within-subject variables. The interaction Repetition X Type of stimuli was not significant ($F(2,12) < 1$) nor the triple interaction Repetition x Type of stimuli x Hemisphere ($F(2,12) = 1.46$, $p = .27$), suggesting a similar repetition suppression effect for the three types of stimuli (figure 1). The significant Repetition by Hemisphere interaction ($F(1,6) = 7.45$, $p = .034$), that was observed can be related to the criterion that led to the selection of the cluster.

Speech vs Music

Table 1 reports the activation for the three conditions relative to the inter-stimulus silent periods. Although the effect size and the extension of the temporal cluster of activation, particularly on the left side, tended to be more important for the mother voice than for the other two stimuli, direct contrasts for speech vs music, mother's vs stranger's voice, and mother's voice vs music, did not reach a significant statistical level.

However, the random analysis of the left-right differences revealed that the responses for these different types of sound were differently lateralized (table 2 and figure 2). A cluster of 60 voxels in the left posterior temporal region was more activated by speech than by music (60 voxels at $p < .01$, $z = 3.40$, cluster-level corrected $p = .002$) while a reverse pattern was observed on the right side (50 voxels at $p < .01$, $z = 3.94$, cluster-level corrected $p = .008$).

For ROI analysis, activation averaged across the voxels of the two clusters observed above was extracted from the original contrast images (mother's voice, stranger's voice and music) and entered in an ANOVA with type of stimuli (3 levels) and hemisphere (2 levels) as within-subject variables. There was a significant left/right asymmetry for both speech conditions (mother's voice: $F(1,6) = 18.52$, $p = .005$; stranger's voice: $F(1,6) = 6.04$, $p = .049$) but not for music $F(1,6) < 1$, figure 2). There was a significant Mother's voice vs Music X Hemisphere interaction ($F(1,6) = 32.74$, $p = .001$), while the stranger's voice vs Music X

hemisphere interaction was only marginally significant ($F(1,6)=4.13$, $p=.088$) and the mother's vs stranger's voice X hemisphere interaction not significant ($F(1,6)=2.16$, $p=.19$).

As seen in figure 2, these two clusters were located over and behind the *planum temporale*. To confirm that this anatomical structure is differently activated by speech and music, we analysed the activations in left and right *planum temporale* as defined from fine-grained anatomical images of another group of 14 infants (64 voxels on the left and 76 voxels on the right side in EPI space; see Methods). An ANOVA with type of stimuli (3 levels) and hemisphere (2 levels) as within-subject variables confirmed the different lateralization for the mother's voice and music responses. There was a significant Mother's voice vs Music X Hemisphere interaction ($F(1,6)=7.71$, $p=.03$), due to a greater left than right response for the mother's voice ($F(1,6)=6.29$, $p=.046$). The other interactions with hemisphere were not significant (stranger's voice vs music X Hemisphere: $F(1,6) < 1$ and mother's vs stranger's voice X Hemisphere $F(1,6) < 1$).

Fixed-effect analyses :

Although the random-effect analysis did not evidence any significant difference in activation to the mother's or to a stranger's voice, infants are known to be able to recognize their mother's voice (DeCasper & Fifer, 1980; Mehler et al., 1978). To further explore the cerebral bases of this capacity, we used a more sensitive fixed-effect analysis which compared the two speech conditions while pooling across the images of all seven subjects. The mother's voice elicited stronger activations than an unknown voice in the anterior prefrontal cortex and the left posterior temporal region, while the reverse comparison revealed a lower level of activation for the mother's voice in the orbito-frontal region, putamen, amygdala. A cluster of the right superior temporal sulcus was also observed (12 vox $z= 5.17$, peak at $x=40$, $y=-6$, $z=-9$) which, however, did not reach significance at the cluster level (table 3). The negative difference between the mother's voice and the other conditions can be essentially explained by a genuine negative BOLD response (deactivation relative to the inter-stimulus silent period) for the mother's voice in this set of regions, rather than an activation for the stranger's voice. Indeed, the BOLD response was always flatter for the stranger's than for the mother's voice (see figure 3). Furthermore, all of these clusters were found significant when a negative contrast for reduced activation to the mother voice relative to silence was computed (table 3 and figure 4).

Discussion

A left-hemisphere advantage for speech processing

Our results indicate that in two-month-old infants, listening to speech already specifically activates the left *planum temporale*. In this area, a left-hemisphere advantage was observed for both speech conditions (mother or stranger voice), while the activations induced by music were symmetrical (figure 2). Structural differences are observed in the posterior part of the temporal region from the last months of gestation (Dubois, Benders et al., 2008; Dubois, Hertz-Pannier et al., 2008; Witelson, 1977). The *planum temporale* is larger on the left side (Chi et al., 1977; Witelson, 1977). The sylvian scissure is steeper and shorter on the right (Sowell et al., 2002), the surface of the right superior temporal sulcus is larger in preterm newborns (Dubois et al., 2007). These structural differences have thus a functional counterpart. However, because of this structural asymmetry, it can be objected that the left and right temporal region that were compared in our asymmetry analysis were not precisely aligned. Note however that we flipped the images by applying a non-linear transformation matrix that optimally aligns the left hemisphere of the infant's template with the right and should therefore have corrected for most of the structural differences. We thus completed the voxel based analysis by an analysis limited to anatomically drawn *planum temporale*. This analysis confirmed that the response to the mother's voice was significantly more left-lateralized than the response to music. This result strengthens our previous fMRI observations of an early left-right functional difference when listening to speech (Dehaene-Lambertz et al., 2002; Dehaene-Lambertz, Hertz-Pannier, Dubois et al., 2006). Crucially, we now show that this functional asymmetry does not extend to all auditory stimuli, and that the left temporal lobe is particularly sensitive to speech, as compared to music, in two-month-old infants.

What factors might explain this left-hemispheric bias for speech stimuli? Music and speech are both structurally organized auditory stimuli, and infants have been shown to be sensitive to the structural organization of both musical and verbal sentences (Hirsh-Pasek et al., 1987; Krumhansl & Jusczyk, 1990). However, these stimuli differ on familiarity as well as on numerous acoustical properties. Among these, the speed of temporal transitions may be an essential factor. An important argument is provided by the observation of a left-hemispheric lateralization to backward speech. Backward speech has been used as a control stimulus for speech because it contains the same fast transitions as forward speech but, contrary to music

stimuli, it violates the suprasegmental phonological properties universally observed in human speech. Backward speech is also highly non-familiar and infants capacities to recognize their native language vanishes when sentences are played backward (Mehler et al., 1988; Ramus, Hauser, Miller, Morris, & Mehler, 2000). Yet, in three-month-olds, no significant difference in lateralization was observed for forward and backward speech in the temporal regions, both inducing stronger responses on the left side in our earlier study (Dehaene-Lambertz, Hertz-Pannier and Dehaene, (2002). In neonates, Pena, Maki et al (2003) reported a significant left-lateralization of activation for forward speech, but not for backward speech, yet the interaction between Hemispheres and type of Speech was not significant. The lack of a significant interaction in these studies might be related to a lack of statistical power, but might also suggest that this region is reacting to fast temporal transitions, present in forward and backward speech but not so massively in music. Note, however, that adults can still correctly identified phonemes much better than chance (72%) in backward words (Binder et al., 2000) leaving open the possibility that the *planum temporale*, which is a region sensitive to phonetic information might be activated by the preserved phonemes in backward speech.

In adults, numerous neuropsychological and neuroimaging studies have demonstrated that the left hemisphere is superior in processing fast temporal changes, while the right hemisphere favours spectral processing (Boemio, Fromm, Braun, & Poeppel, 2005; Schwartz & Tallal, 1980; Zatorre & Belin, 2001). However, the left-hemispheric lateralization of phonetic processing goes beyond acoustical characteristics, suggesting that constraints other than low-level perceptual processes are driving lateralization in adults. For example, for identical stimuli, the left-hemispheric lateralization is increased when the stimuli are perceived as linguistic productions in the listener's native language (Celsis et al., 1999; Dehaene-Lambertz et al., 2005; Gandour et al., 2002; Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003; Shtyrov, Pihko, & Pulvermuller, 2005). In two-month-old infants, Bristow et al (2009) reported that the sources of the mismatch responses were left lateralized for a vowel mismatch and right lateralized for a gender mismatch. Because both vowel identification and voice categorization rely largely on the analysis of spectral content, the different lateralization of the sources suggests that, as in adults, functional lateralization in infants cannot be explained only by the physical features of the stimuli and is driven by their functional contents (e.g. the conveying of linguistic features versus non-linguistic identity information).

Future studies should further explore the origin of the greater involvement of the left *planum temporale* in speech processing. Speculatively, it seems possible that language

lateralization starts as a minimal, domain-general genetic bias of the left-hemisphere for processing rapidly changing stimuli—a bias that would rapidly be extended through learning to other properties of the speech signal (i.e. vowel analysis), eventually favouring the rapid development of a left-hemispheric specialization for essentially all levels of speech processing in our species. Alternatively, it is possible that several simultaneous biases, at multiple phonological, lexical and syntactic levels, jointly contribute to language lateralization. Recently developed methods of “co-lateralization” analyses in large groups of subjects (Pinel & Dehaene, 2009), together with fMRI-genetics correlation, may ultimately resolve this issue.

Repetition effect

A repetition suppression effect was observed in the left temporal region, and was similar for all three conditions (sentences and music). Repetition suppression is classically observed when the same representation is accessed multiples times (Naccache & Dehaene, 2001). Here the same 4 s segment was repeated three times, implying that the regions that were sensitive to repetition were integrating information across a rather long time-window of four seconds. In adults, when the same sentence is repeated at a 14-s interval, repetition suppression is observed in the superior temporal sulcus, whereas more dorsal regions in Heschl’s gyrus and superior temporal gyrus remain unaffected by repetition, presumably because they are only sensitive to recent and local temporal features (Dehaene-Lambertz, Dehaene et al., 2006). It is striking that the present results replicate this pattern in two-month infants: as seen in figure 1, it is also the most ventral part of the activation induced by the auditory stimuli which demonstrated a repetition suppression effect. This dorsal-ventral difference confirms our hypothesis that the temporal regions are hierarchically organized from the beginning of life, and fuels the speculation that the superior temporal sulcus may contain regions sensitive to higher-level and temporally more extended units than Heschl’s gyrus and its immediate vicinity (Dehaene-Lambertz, Hertz-Pannier, Dubois et al., 2006)

In our previous study in infants (Dehaene-Lambertz, Dehaene et al., 2006), using a slow event design with one sentence every 14 s, we did not observe repetition suppression in any regions. It is hard to nail down the exact reason for this null effect: the high variance of infants’ data due to their variable commitment in the task, their various stages of brain maturation, or the residual artefacts due to movements can always be suspected to have induced a lack of statistical power. However, it is also possible that the difference between the two sets of data is due to a key distinction in experimental design: the immediate repetition of

sentences in the present experiment, versus delayed repetition in our previous work. This hypothesis would suggest that the memory buffer of these regions at this age might not be able to hold information throughout 14 s in order to detect the similarity between two consecutive sentences. It is also noticeable that only a cluster in the left temporal region was detected. Although there was no significant difference in the lateralization of this effect in the random-effect analysis, this result might suggest another difference between the two hemispheres regarding their capacity to compute and maintain long and abstract units. In the future, parametric studies of fMRI repetition suppression, systematically varying the time interval between consecutive sentence repetitions, might provide a methodology to evaluate the duration of the different buffers involved in memory for speech at different levels of processing and in both hemispheres.

Mother's voice is special

Just after birth, behavioural studies indicate that infants already recognize their mother's voice (DeCasper & Fifer, 1980; Mehler et al., 1978). This competence must therefore be based, at least in part, on a sense of familiarity acquired during the last weeks of gestation. The mother's voice may affect infants' brain responses through different mechanisms. Her voice is a familiar stimulus that might trigger different memory components (episodic, semantic, emotional, etc..) and elicit associations with faces and emotions previously experienced in association with this voice. Because it is usually associated with feeding and soothing, the mother's voice can induce positive emotional responses. Finally, because it is a salient stimulus in the infant's environment, it can also elicit attention and facilitate speech processing. It is difficult to separate the contribution of these different factors, which might have been different from one baby to the next depending on his/her vigilance as well as his/her previous experience. Although the response was more extended for the mother voice than for an unknown voice, the random-effect analyses did not reveal significant differences at the considered statistical level. The null effect in the random analyses can be due to the small number of infants and to the fact that most of them were asleep or falling asleep during the experiment, therefore reducing the strength of fMRI responses to external stimuli (Dehaene-Lambertz et al., 2002; Portas et al., 2000). The fixed effect analysis uncovered differences in activation between the two voice conditions in our population. Significantly stronger responses were elicited by the mother's voice, leading to significant differences in the left posterior part of the temporal lobe and in the left and right

anterior prefrontal cortex. Furthermore, the reverse comparison isolated several regions involved in emotional processing in adults such as the orbito frontal cortex, the putamen and the amygdala which seemed to be de-activated by the mother's voice. Although these results were obtained by fixed effect analysis in a small group of infants, a statistical technique that does not guarantee generalization to a new group of participants, and will thus need to be confirmed and replicated, they point to a potentially interesting linkage of linguistic and emotional network in infants that might play an important role in learning.

Among the two regions whose activity was enhanced by the mother's voice (figure 3 and 4), the left posterior temporal region raises the question of the impact of the mother's voice on linguistic processing. The observed effect might be attributed to an increase in overall attention, stimulated by this familiar positive stimulus, which might have led to a global enhancement of brain activity. However it is noticeable that this response enhancement by the mother's voice was observed only in the most posterior region of the temporal lobe, not in more anterior and superior primary/secondary auditory regions. The posterior temporal region is involved in phonological representations in adults (Caplan, Gow, & Makris, 1995) and is thought to be a plausible cortical source of phonetic mismatch responses that can be recorded with scalp event-related potentials in infants (Bristow et al., in press; Dehaene-Lambertz & Baillet, 1998). It may therefore seem surprising that a phonological region would be so sensitive to a specific voice. Because the speaker's characteristics have a high impact on the acoustic realization of speech, some authors postulate that phonetic processing involves a normalization procedure to neutralize non pertinent acoustical variations, including those induced by different speaker voices (Blumstein & Stevens, 1981). Others, however, have proposed that speaker information is encoded along with phonetic information (Goldinger, 1996). As in adults (Dahan, Drucker, & Scarborough, 2008), infant studies bring support for both hypotheses. On the one hand, infants can notice that the same phoneme is repeated even when it is produced by different speakers (Dehaene-Lambertz & Pena, 2001; P.K. Kuhl, 1983) suggesting that they have access to an abstract, speaker-invariant representation of phonemes. On the other hand, they have difficulties in memorizing phonological details when syllables are produced by many talkers (Jusczyk, Pisoni, & Mullennix, 1992) or in recognizing previously learned words in passages produced by a different talker (Houston & Jusczyk, 2003). These results suggest either that they have difficulties to access an abstract speaker invariant representation in the absence of convergent contextual information, or that only instances retaining contextual particularities are stored at this age.

With respect to this long-standing debate, the present results cannot disentangle whether normalization stage is facilitated when the voice has been encountered many times (perceptual adaptation) or whether infants have stored parents' productions as instances. They nevertheless suggest that phonetic processing in infants is sensitive to talker characteristics and can be improved by listening to a highly familiar voice such as the mother's voice. This result is congruent with a behavioural study showing that, in the presence of distracting background speech, infants are better at learning words when they are spoken by the mother rather by an unfamiliar speaker (Barker & Newman, 2004). An ERP study comparing the response to a word pronounced by the mother and by an unfamiliar voice also showed that the early auditory components were accelerated for the mother's voice (Purhonen, Kilpelainen-Lees, Valkonen-Korhonen, Karhu, & Lehtonen, 2004). Our finding can also explain why the clarity of the mothers' speech has a strong impact on infants' phoneme discrimination capacities (Liu, Kuhl, & Tsao, 2003).

A second cluster of increased activation to the mother's voice was located in the anterior prefrontal cortex. In adults, prefrontal cortex has been found to react to the call of the participant's name, even in sleeping subjects (Portas et al., 2000). Grossman et al (2008) reported increased brain activity over frontal areas when four-month-old infants perceived communication cues, both as measured by near-infrared spectroscopy and by EEG recordings (increase in the amplitude of gamma-band activity). Minagawa-Kawai et al (2009) reported a significant difference in OxyHb when one-year old infants were looking at their smiling mother relative to an unknown face. Thus, this area may be important for mother-infant contact and emotional attachment.

By contrast, a set of brain areas in the subcortical regions, medial temporal lobe and orbitofrontal cortex were more activated by the unknown voice. These areas are all involved in emotional processing in adults (Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). The difference between mother's and stranger's voice appears to be related to a "deactivation" or "negative BOLD response" to the mother's voice, rather than to a genuine activation to the stranger's voice (table 3, figure 3 and 4). Indeed most of these clusters were also significant when testing for a reduction in the BOLD response to the mother's voice. Because negative BOLD responses can be produced by different causes in infants, it is difficult to have a clear understanding of this result. Negative BOLD responses have been observed mainly in visual

areas during visual stimulation (Altman & Bernal, 2001; Morita et al., 2000; Yamada et al., 1997), and have been related to an immaturity of infants' hemodynamic coupling, possibly associated with a higher demand in oxygen due to rapid synaptogenesis. A higher extraction of oxygen would produce higher concentrations of deoxyhemoglobin which would not be adequately compensated by the increase of cerebral blood flow when stimulation occurs (Morita et al., 2000). Like visual regions, the central regions (basal ganglia, thalamus, medial temporal lobe) show a rapid increase in metabolism during the first weeks of life (Chugani, Phelps, & Maziotta, 1987) that might induce negative BOLD responses when they are stimulated. An alternative interpretation is that the observed deactivation relates to a higher activation of these regions when infants did not hear their mother, and a genuine decrease of neuronal activity (soothing effect) when the mother's voice was perceived. Indeed, a great variety of activation studies in animals and adult humans associate the amygdala with negative emotions, which might be alleviated when the baby is distracted by suddenly hearing his or her mother. However, it should be remembered that several brain imaging and neuropsychological studies also support a role of the amygdala in positive emotions processing (Fecteau, Belin, Joanne, & Armony, 2007; Lanteaume et al., 2007). It has therefore been proposed that the amygdala, which comprises several sub-nuclei, operates overall as a detector of biologically relevant stimuli not limited to fearful responses (Sander, Grafman, & Zalla, 2003). In parallel with a rapid and automatic response in the amygdala, emotional information conveyed by voice may target the orbito-frontal cortex when explicit emotional judgements are requested in adults (Sander et al., 2005; Wildgruber et al., 2006). Our results suggest that both pathways are already available to infants.

It is noteworthy that we observed opposite effects in anterior prefrontal versus orbito-frontal cortices, these regions showing inverse responses to the known and unknown voices (figure 4). This might remind the spatial separation observed in adults' frontal area between activity elicited when thinking about the self and familiar other versus unknown others (Amodio & Frith, 2006), or when focusing toward external world vs internal state (Wicker, Ruby, Royet, & Fonlupt, 2003). Although frontal cortex has long been considered as inactive in infants because its maturation is delayed and protracted, evidence is accumulating revealing its involvement in early cognitive functions (Dehaene-Lambertz, Dehaene et al., 2006; Dehaene-Lambertz et al., 2002; Fransson et al., 2009; Fransson et al., 2007; Gao et al., 2009; Grossmann et al., 2008). One of the strongest evidence of an early functional frontal cortex comes from Fransson et al's studies (2007 and 2009). Beside the expected networks in the

sensory-motor regions (primary visual cortex, bilateral sensorimotor areas, bilateral auditory cortex), these authors reported a spontaneous synchronised activity in the medial and dorsolateral prefrontal areas in sleeping newborns. The frontal network represented one of the five resting state networks present at this age (the last one involved the precuneus area, the lateral parietal cortex, and the cerebellum). This suggests a functional connectivity already efficient within frontal areas as it is the case in the in the more mature sensory-motor regions.

Finally, we were expecting a difference between both voices in the anterior part of the right STS because this region might represent a specific voice area as proposed by Belin et coll (2004). A high local z-value was indeed observed for coordinates in that region (table 3), but the cluster was not significant when corrected at the cluster-level. In adults, this region was detected in speech-non-speech comparisons, voice discrimination, and emotional prosody categorisation tasks (Belin et al., 2004; Sander et al., 2005). Here, our comparison contrasted two voices that were both highly intonated (motherese), with positive emotional valence, and in an experimental paradigm that did not require voice discrimination. Thus, the differences between the two voices might have been too weak to induce a strong difference in this region. It is also noticeable that this region did not show up when music and speech were compared, which might suggest a loss of fMRI sensitivity in this region where susceptibility artifacts can be a problem.

Conclusion

Infant brain imaging studies are still in their infancy. The difficulties of fMRI at this age, combined with the neuronal and vascular immaturity and the small size of the targeted brain structures, present serious obstacles to this research. Nevertheless a small but growing infant neuroimaging literature points to the existence, in the first few months of life, of a well-structured cortical organization. In the superior temporal regions, we observed a dorsal-ventral gradient in the sensitivity to stimulus repetition and a left-right difference in the sensitivity to speech stimuli. How this particular organization supports speech acquisition is still an open question but a research program looking for genetic variants specifically expressed in the posterior temporal region in the human lineage might be a productive endeavour. Acknowledging the existence of strong genetic constraints on the organization of the perisylvian regions does not preclude environmental influences. The present results show clearly that learning also plays a major role in structuring the infant's brain networks,

inasmuch as the mother's voice has a strong impact on several brain regions involved in emotion and communication, but also on the left-hemispheric language network, particularly in the posterior temporal region.

Acknowledgments:

This study was supported by INSERM, CEA, the McDonnell foundation, IFR49, APETREIMC, "the European NEST-PATHFINDER program "Neurocom" (Contract N°12738)", and the Agence Nationale pour la Recherche (project NIBB). We thank Thierry Brunet (Ravier-Touzard) for designing a baby chair specifically adapted to the head coil and Pr Le Bihan for his support.

References:

- Altman, N. R., & Bernal, B. (2001). Brain activation in sedated children: auditory and visual functional MR imaging. *Radiology*, *221*(1), 56-63.
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*, *7*(4), 268-277.
- Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2), B45-53.
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci*, *8*(3), 129-135.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309-312.
- Bertoncini, J., Morais, J., Bijeljic-Babic, R., McAdams, S., Peretz, I., & Mehler, J. (1989). Dichotic perception and laterality in neonates. *Brain and Language*, *37*, 591-605.
- Best, C. T., Hoffman, H., & Glanville, B. B. (1982). Development of infant ear asymmetries for speech and music. *Perception and Psychophysics*, *31*, 75-85.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and non speech sounds. *Cerebral Cortex*, *10*(5), 512-528.
- Blumstein, S. E., & Stevens, K. N. (1981). Phonetic features and acoustic invariance in speech. *Cognition*, *10*, 25-32.
- Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci*, *8*(3), 389-395.
- Bowers, D., Coslett, H. B., Bauer, R. M., Speedie, L. J., & Heilman, K. M. (1987). Comprehension of emotional prosody following unilateral hemispheric lesions: Processing defect versus distraction defect. *Neuropsychologia*, *25*, 317-328.
- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., et al. (in press). Hearing faces: Crossmodal representations of speech in two-month-old infants. *Journal of Cognitive Neuroscience*.
- Caplan, D., Gow, D., & Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology*, *45*(2), 293-298.
- Celsis, P., Doyon, B., Boulanouar, K., Pastor, J., Demonet, J. F., & Nespoulous, J. L. (1999). ERP correlates of phoneme perception in speech and sound contexts. *Neuroreport*, *10*(7), 1523-1527.
- Chi, J. G., Dooling, E. C., & Gilles, F. H. (1977). Gyral development of the human brain. *Annals of Neurology*, *1*, 86-93.
- Chugani, H. T., Phelps, M. E., & Maziotta, J. C. (1987). Positron emission tomography study of human brain functional development. *Annals of Neurology*, *22*, 487-497.
- Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: adjusting the signal or the representations? *Cognition*, *108*(3), 710-718.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mother's voices. *Science*, *208*, 1174-1176.
- Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience*, *12*(3), 449-460.
- Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *NeuroReport*, *9*, 1885-1888.
- Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, *370*, 292-295.
- Dehaene-Lambertz, G., Dehaene, S., Anton, J. L., Campagne, A., Ciuciu, P., Dehaene, G. P., et al. (2006). Functional segregation of cortical language areas by sentence repetition. *Hum Brain Mapp*, *27*(5), 360-371.

- Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298, 2013-2015.
- Dehaene-Lambertz, G., Hertz-Pannier, L., & Dubois, J. (2006). Nature and nurture in language acquisition: anatomical and functional brain-imaging studies in infants. *Trends Neurosci*, 29(7), 367-373.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., & Dehaene, S. (2008). How does early brain organization promote language acquisition in humans? *European Review*, 16(4), 399-411.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M., et al. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc Natl Acad Sci U S A*, 103(38), 14240-14245.
- Dehaene-Lambertz, G., Pallier, C., Serniklaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural Correlates of Switching from Auditory to Speech Perception. *NeuroImage*, 24, 21-33.
- Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport*, 12, 3155-3158.
- Dubois, J., Benders, M., Borradori-Tolsa, C., Cachia, A., Lazeyras, F., Ha-Vinh Leuchter, R., et al. (2008). Primary cortical folding in the human newborn: an early marker of later functional development. *Brain*, 131(Pt 8), 2028-2041.
- Dubois, J., Hertz-Pannier, L., Cachia, A., Mangin, J. F., Le Bihan, D., & Dehaene-Lambertz, G. (2008). Structural Asymmetries in the Infant Language and Sensori-Motor Networks. *Cereb Cortex*.
- Entus, A. K. (1977). Hemispheric asymmetry in processing of dichotically presented speech and nonspeech stimuli by infants. In S. J. Segalowitz & F. A. Gruber (Eds.), *Language Development and Neurological Theory* (pp. 63-73). New York: Academic Press.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., et al. (2006). Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage*, 30(2), 580-587.
- Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage*, 36(2), 480-487.
- Fransson, P., Skiold, B., Engstrom, M., Hallberg, B., Mosskin, M., Aden, U., et al. (2009). Spontaneous brain activity in the newborn brain during natural sleep - An fMRI study in infants born at full term. *Pediatr Res*.
- Fransson, P., Skiold, B., Horsch, S., Nordell, A., Blennow, M., Lagercrantz, H., et al. (2007). Resting-state networks in the infant brain. *Proc Natl Acad Sci U S A*, 104(39), 15531-15536.
- Gandour, J., Wong, D., Lowe, M., Dziedzic, M., Sathamnuwong, N., Tong, Y., et al. (2002). A cross-linguistic FMRI study of spectral and temporal cues underlying phonological processing. *J Cogn Neurosci*, 14(7), 1076-1087.
- Gao, W., Zhu, H., Giovanello, K. S., Smith, J. K., Shen, D., Gilmore, J. H., et al. (2009). Evidence on the emergence of the brain's default network from 2-week-old to 2-year-old healthy pediatric subjects. *Proc Natl Acad Sci U S A*, 106(16), 6790-6795.
- Glanville, B. B., Best, C. T., & Levenson, R. (1977). A cardiac measure of cerebral asymmetries in infant auditory perception. *Developmental Psychology*, 13, 54-59.
- Glaser, H. (2007). *Les aires du langage chez le bebe: Analyse morphometrique des sillons*. Pierre Mendes France, Grenoble II.
- Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *J Exp Psychol Learn Mem Cogn*, 22(5), 1166-1183.
- Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)*, 107(1-3), 293-321.

- Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., et al. (2008). Early cortical specialization for face-to-face communication in human infants. *Proc Biol Sci*, 275(1653), 2803-2811.
- Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269-286.
- Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neurosci Res*, 54(4), 276-280.
- Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *J Exp Psychol Hum Percept Perform*, 29(6), 1143-1154.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological Grammar Shapes the Auditory Cortex: A Functional Magnetic Resonance Imaging Study. *Journal of Neuroscience*, 23, 9541 - 9546.
- Jusczyk, P. W. (1997). *The discovery of spoken language*: MIT Press.
- Jusczyk, P. W., Pisoni, D. B., & Mullenix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, 43(3), 253-291.
- Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science*, 1, 70-73.
- Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, 6, 263-285.
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat Rev Neurosci*, 5(11), 831-843.
- Lanteaume, L., Khalfa, S., Regis, J., Marquis, P., Chauvel, P., & Bartolomei, F. (2007). Emotion induction after direct intracerebral stimulations of human amygdala. *Cereb Cortex*, 17(6), 1307-1313.
- Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mother's speech clarity and infants' speech discrimination skills. *Developmental Science*, 6, F1-F10.
- Mehler, J., Bertoncini, J., & Barriere, M. (1978). Infant recognition of mother's voice. *Perception*, 7(5), 491-497.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.
- Minagawa-Kawai, Y., Matsuoka, S., Dan, I., Naoi, N., Nakamura, K., & Kojima, S. (2009). Prefrontal Activation Associated with Social Attachment: Facial-Emotion Recognition in Mothers and Infants. *Cereb Cortex*, 19, 284-292.
- Morita, T., Kochiyama, T., Yamada, H., Konishi, Y., Yonekura, Y., Matsumura, M., et al. (2000). Difference in the metabolic response to photic stimulation of the lateral geniculate nucleus and the primary visual cortex of infants: a fMRI study. *Neurosci Res*, 38(1), 63-70.
- Naccache, L., & Dehaene, S. (2001). The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex*, 11(10), 966-974.
- Pena, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., et al. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proc Natl Acad Sci U S A*, 100(20), 11702-11705.
- Pinel, P., & Dehaene, S. (2009). Beyond Hemispheric Dominance: Brain Regions Underlying the Joint Lateralization of Language and Arithmetic to the Left Hemisphere. *J Cogn Neurosci*.
- Portas, C. M., Krakow, K., Allen, P., Josephs, O., Armony, J. L., & Frith, C. D. (2000). Auditory processing across the sleep-wake cycle: simultaneous EEG and fMRI monitoring in humans. *Neuron*, 28(3), 991-999.

- Purhonen, M., Kilpelainen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2004). Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *Int J Psychophysiol*, *52*(3), 257-266.
- Ramus, F., Hauser, M., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*, 349-351.
- Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Rev Neurosci*, *14*(4), 303-316.
- Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *Neuroimage*, *28*(4), 848-858.
- Schwartz, J., & Tallal, P. (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science*, *207*(4437), 1380-1381.
- Shtyrov, Y., Pihko, E., & Pulvermuller, F. (2005). Determinants of dominance: is language laterality explained by physical or linguistic features of speech? *Neuroimage*, *27*(1), 37-47.
- Sowell, E. R., Thompson, P. M., Rex, D., Kornsand, D., Tessner, K. D., Jernigan, T. L., et al. (2002). Mapping sulcal pattern asymmetry and local cortical surface gray matter distribution in vivo: maturation in perisylvian cortices. *Cerebral Cortex*, *12*(1), 17-26.
- Sun, T., Collura, R. V., Ruvolo, M., & Walsh, C. A. (2006). Genomic and evolutionary analyses of asymmetrically expressed genes in human fetal left and right cerebral cortex. *Cereb Cortex*, *16 Suppl 1*, i18-25.
- Tzourio, N., Nkanga-Ngila, B., & Mazoyer, B. (1998). Left planum temporale surface correlates with functional dominance during story listening. *Neuroreport*, *9*(5), 829-833.
- Van Essen, D. C. (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage*, *28*(3), 635-662.
- Vargha-Khadem, F., & Corballis, M. (1979). Cerebral asymmetry in infants. *Brain and Language*, *8*, 1-9.
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental model of speech processing. *Language Learning and Development*, *1*(2), 197-234.
- Wicker, B., Ruby, P., Royet, J. P., & Fonlupt, P. (2003). A relation between rest and the self in the brain? *Brain Res Brain Res Rev*, *43*(2), 224-230.
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog Brain Res*, *156*, 249-268.
- Witelson, S. F. (1977). Anatomic asymmetry in the temporal lobes: its documentation, phylogenesis, and relationship to functional asymmetry. *Ann N Y Acad Sci*, *299*, 328-354.
- Yamada, H., Sadato, N., Konishi, Y., Kimura, K., Tanaka, M., Yonekura, Y., et al. (1997). A rapid brain metabolic change in infants detected by fMRI. *NeuroReport*, *8*, 3775-3778.
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*(10), 946-953.

Table 1: Random effect analyses

Analysis	Area	Infant template coordinates			SPM t-tests		
		x	y	z	No. of voxels in cluster	Cluster-level P value (corrected)	Z value at local maximum
Main effect of sound	Left STG/STS	-37	-17	6	165	<0.001	3.61
		-40	-6	-3			3.10
	Left Planum temporale	-31	-26	6	81	0.009	3.61
	Right STG	40	-14	3			4.26
		46	-17	3			3.34
	Right Heschl's gyrus	46	-11	0			3.26
Repetition effect	Left STG/STS	-40	-11	0	88	0.005	4.15
		-43	-23	-6			3.77
		-46	-3	-9			2.78
Mother Voice	Left STS/STG	-34	-17	9	210	<0.001	4.13
		-40	-9	-3			3.33
	Left Planum temporale	-29	-29	3			3.52
	Left Precentral gyrus	-29	-20	23			3.55
	Right STG	40	-14	3	52	0.073	4.37
Foreign Voice	Left STG	-34	-23	6	24	0.787	3.31
	Right STG/STS	43	0	6	59	0.046	3.19
Music	Left STG	-31	-26	6	82	0.011	3.55
	Right Heschl's gyrus	34	-20	6	62	0.05	3.83
	Right STG	46	-11	0			3.36

STS: Superior temporal sulcus, STG: Superior temporal gyrus

Table 2: Random-effect analyses of asymmetry

Analysis	Area	Infant template coordinates			SPM t-tests		
		x	y	z	No. of voxels in cluster	Cluster-level P value (corrected)	Z value at local maximum
Speech > Music	Left Planum temporale and posterior temporal region (L > R)	-20	-31	17	60	0.002	3.40
		-34	-37	11			2.67
Music > Speech	Right Planum temporale (R > L)	23	-34	17	50	0.008	3.94
		37	-31	0			2.55
Mother's voice	Left Planum temporale (L > R)	-31	-34	9	45	0.013	3.05
		-31	-20	11			2.70
Stranger's voice	Left Planum temporale (L > R)	-20	-37	20	4	NS	3.23
		-23	-31	11	5	NS	3.06
	Right occipital area (R>L)	-29	-57	-17	30	0.053	3.38
		-26	-49	-20			2.92
Music	No significant voxel						

Table 3: Fixed effect analyses

Mother's > Stranger's Voice	Right Anterior Prefrontal Cortex	11	43	11	157	<0.001	5.84
	Left Anterior Prefrontal Cortex	-9	34	20	87	<0.001	5.22
	Left posterior temporal region	-40	-37	0	69	<0.001	4.66
Stranger's > Mother's Voice	Right and Left orbito-frontal regions	6	46	-3	51	<0.001	4.93
		-17	31	-9			3.91
	Right Putamen-Amygdala	17	0	-6	39	<0.001	4.93
	Left Amygdala	-14	-6	-9	24	0.002	4.49
	Brainstem	9	-17	-11	19	0.007	4.20
	Right anterior STS and STG	40	-5.7	-8.6	12	0.141	5.17
	- Mother's voice (Positive Bold Response)	Left STG,STS, Planum temporale	-37	-17	2.9	228	<0.001
Left precentral gyrus		-17	-29	40	52	<0.001	5.61
Right STG and STS		46	-17	-2.9	65	<0.001	5.59
Right Anterior Prefrontal Cortex		20	40	2.9	40	<0.001	4.78
Left Anterior Prefrontal Cortex		-29	20	8.6	37	0.001	4.20
Left Anterior Prefrontal Cortex		-20	37	5.7	47	<0.001	4.12
Supplementary motor area		-11	-8.6	37	23	<0.001	3.94
- Mother's voice (Negative Bold Response)	Right lateral occipital sulcus	29	-57	11	39	0.001	5.12
	Right Insula/Amygdala/putamen	31	8.6	-5.7	44	<0.001	4.98
	Right STS	34	-20	-11	19	0.028	4.13
- Stranger's voice (Positive Bold Response)	Right STG and STS	40	-14	-2.9	47	<0.001	5.46
	Right-Left orbito-frontal mesial cortic	11	49	-11	90	<0.001	5.36
	Left STG and STS	-34	-20	2.9	61	<0.001	5.06
- Stranger's voice (Negative Bold Response)	no suprathreshold cluster						

Figure legends

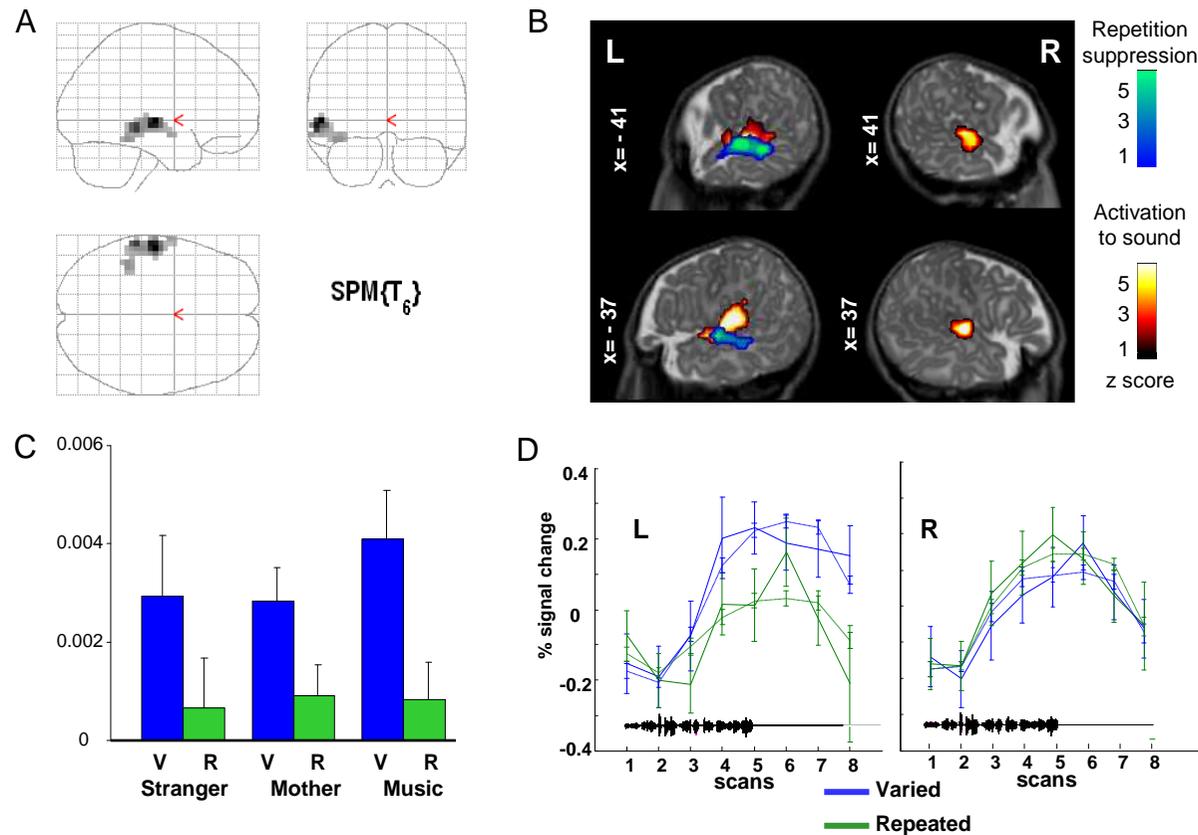
Figure 1 : **Repetition suppression effect.** A cluster along the left superior temporal sulcus showed a reduced BOLD response when the same 4-second segment was repeated, whether it consisted of speech or music. A) Glass brain and B) projection of the cluster (cold scale) on 2D sagittal slices of a three-month-old T2 anatomical image. Activations for sound are projected using a hot scale C) Boxplot of the individual activations averaged over this cluster for each of the auditory stimulus and for varied (V) and repeated (R) segments (arbitrary units). D) Mean BOLD response for repeated and not-repeated segments at the peak maximum of the effect $x=-40$ $y=-11$ $z=0$ and at its right symmetric voxel. Note that the coordinates are provided relative to our infant brain template. They can be roughly compared to adult MNI coordinates after multiplication by a factor of 1.4.

Figure 2: **Hemispheric asymmetries in the functional activations to speech and music in the infant *planum temporale*.** A) Asymmetric clusters of activation of the speech vs Music comparisons are projected on sagittal slices of a three-month-old T2 anatomical image. On both sides, the clusters are located over and behind the *planum temporale*. B) Boxplot of the individual activations averaged over the left and right clusters for each of the auditory stimulus (arbitrary units). There is a significant left/right asymmetry for both speech conditions but not for music.

Figure 3: **Mother's vs Stranger's voice:** Areas of significantly increased (red–yellow scale) and decreased (blue–cyan scale) BOLD response when listening to the mother's voice relative to a stranger's voice. The activations are projected over sagittal slices, spaced every 3 mm from the left to the right side. The BOLD response at major maxima is presented for both conditions (see table 3 for coordinates).

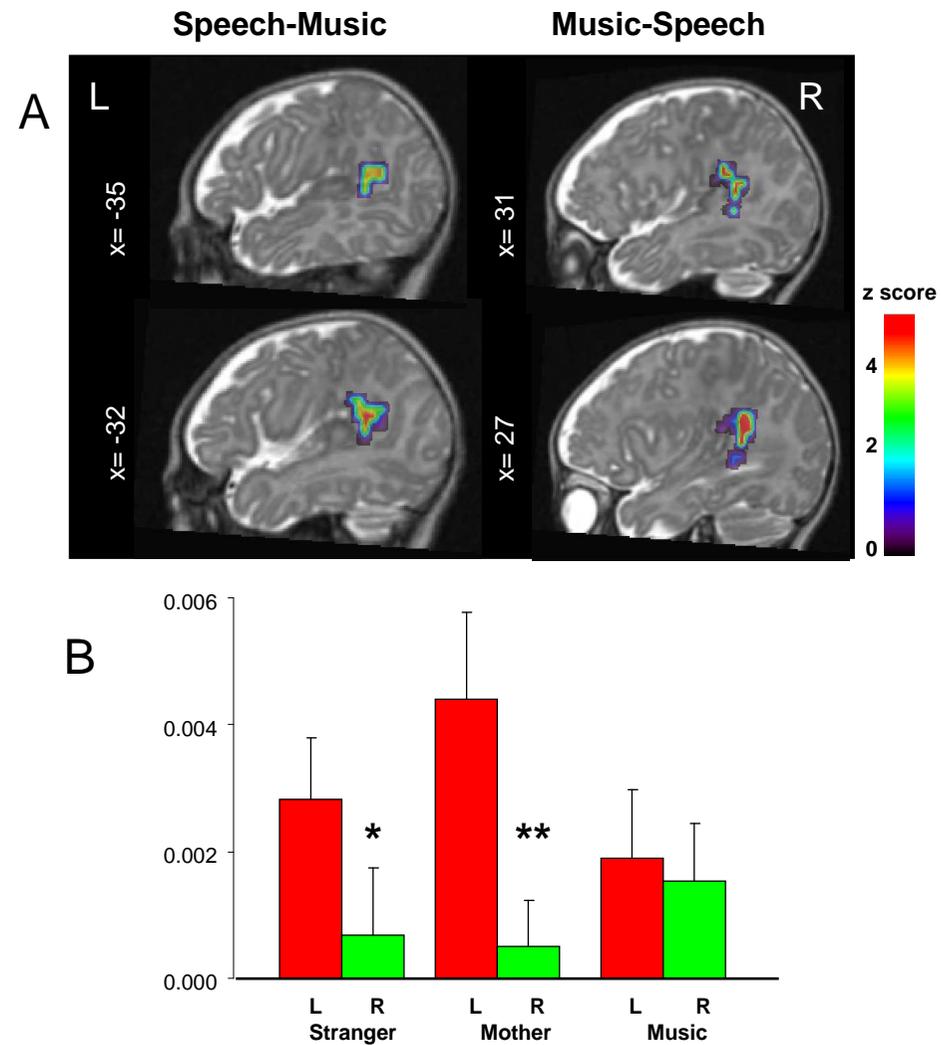
Figure 4: **Mother's vs Stranger's voice:** Areas of significantly increased and decreased BOLD response in a fixed-effect analysis presented on a transparent brain when infants listened to the mother's voice and to a stranger's voice relative to silence.

Figure 1



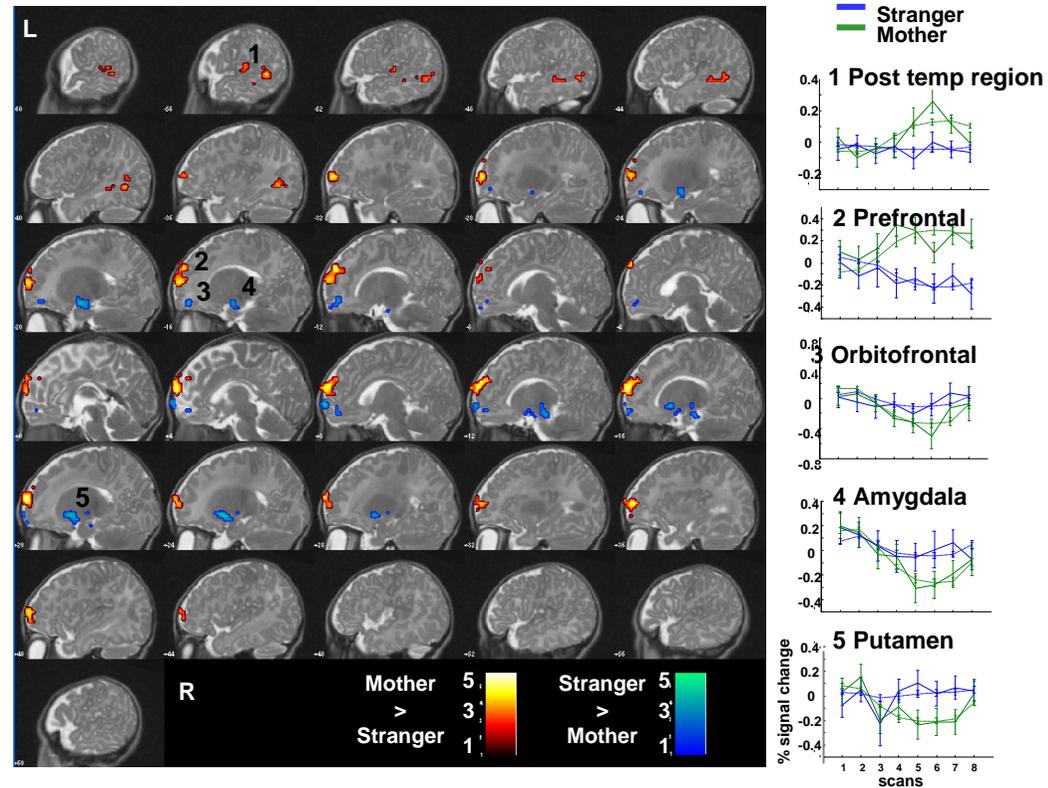
Repetition suppression effect. A left cluster along the superior temporal sulcus is sensitive to repetition of the same 4s segment in all three conditions. A) Glass brain and B) projection of the cluster (cold scale) on 2D sagittal slices of a three-month-old T2 anatomical image. Activations for sound are projected using a hot scale C) Boxplot of the individual activations averaged over this cluster for each of the auditory stimulus and for varied (V) and repeated (R) segments (arbitrary units). D) Mean BOLD response for repeated and not-repeated segments at the peak maximum of the effect $x=-40$ $y=-11$ $z=0$ and at its right symmetric voxel. Note that the coordinates are provided relative to our infant brain template. They can be roughly compared to adult MNI coordinates after multiplication by a factor of 1.4.

Figure 2



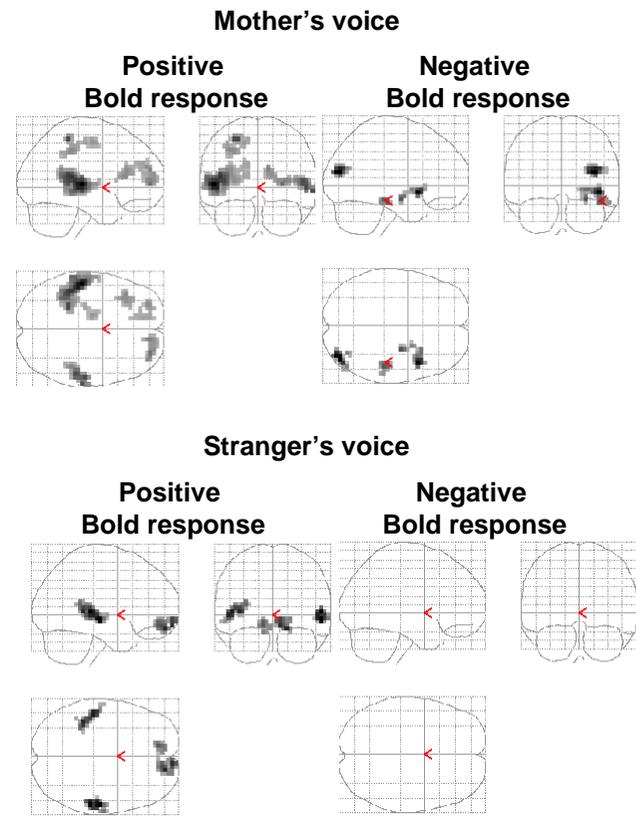
Functional asymmetries for speech and music in the infant *planum temporale*. A) Asymmetric clusters of activation of the speech vs Music comparisons are projected on sagittal slices of a three-month-old T2 anatomical image. On both sides, the clusters are located over and behind the *planum temporale*. B) Boxplot of the individual activations averaged over the left and right clusters for each of the auditory stimulus (arbitrary units). There is a significant left/right asymmetry for both speech conditions (mother's voice: $F(1,6)=18.52$, $p=.005$; stranger's voice: $F(1,6)=6.04$, $p=.049$) but not for music $F(1,6) < 1$.

Figure 3



Mother's vs Stranger's voice: Areas of significantly increased (red–yellow scale) and decreased (blue–cyan scale) MR signal intensity when listening to the mother's voice relative to a stranger's voice. The activations are projected over sagittal slices, spaced every 3 mm from the left to the right side. The BOLD response at major maxima is presented for both conditions (see table 3 for coordinates).

Figure 4



Mother's vs Stranger's voice: Areas of significantly increased and decreased MR signal intensity in a fixed-effect analysis presented on a transparent brain when infants listen to the mother's voice and to a stranger's voice relative to silence.