Current Biology, Volume *21*

Supplemental Information

Early Specialization for Voice and Emotion Processing in the Infant Brain

Anna Blasi, Evelyne Mercure, Sarah Lloyd-Fox, Alex Thomson, Michael Brammer, Disa Sauter, Quinton Deeley, Gareth J. Barker, Ville Renvall, Sean Deoni, David Gasston, Steven C.R. Williams, Mark H. Johnson, Andrew Simmons, and Declan G.M. Murphy

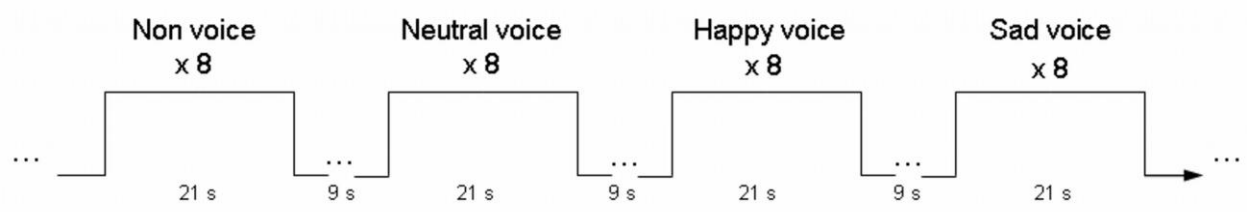

Figure S1. Stimulus Presentation
The stimuli were organized in a block design, with a presentation time of 21 seconds and a rest period of 9 seconds between consecutive sounds. Each condition was presented 8 times, with the *nonvoice* and *neutral voice* conditions given more weight at the beginning of the sequence.

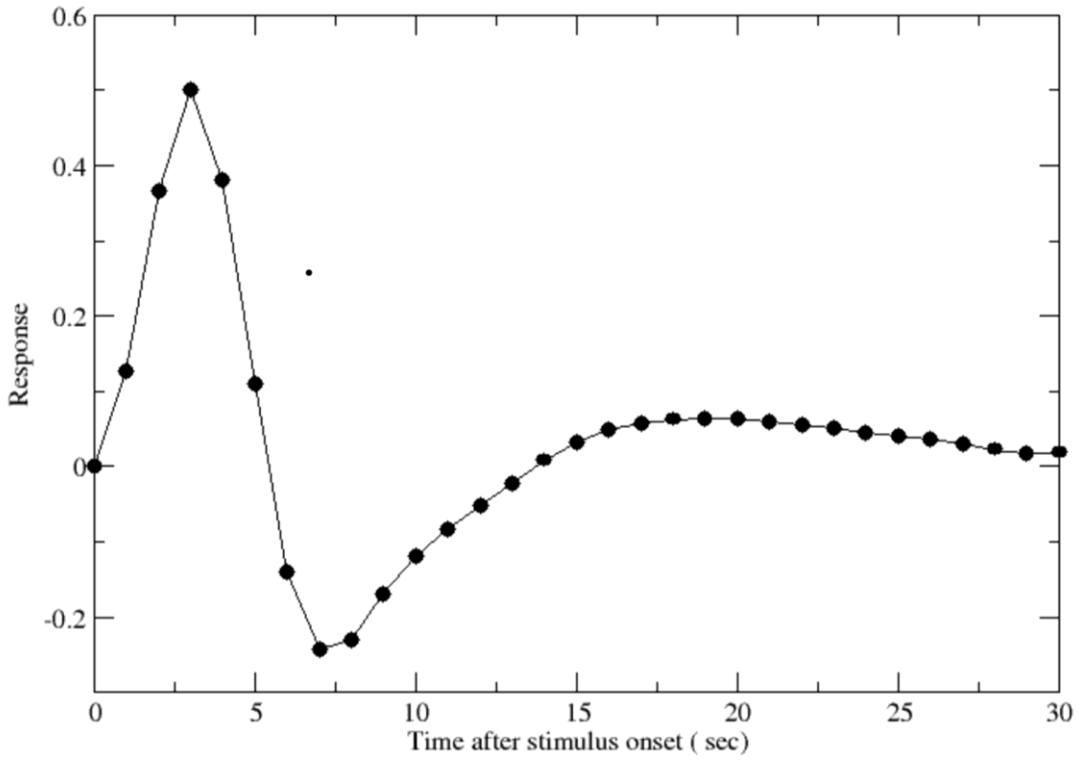

Figure S2. Example of Infant BOLD Response (Group Average)
The response in infants was faster than what would be expected in adults, with a peak at ~ 3 sec after stimulus onset and a significant undershoot with a minimum at ~6 sec (see [S1] for the changes in the BOLD response in children and adults).

Table S1. All Sounds versus Rest Contrast, Related to Figure 1

		Infant template Talairach coordinates			Activation clusters		
	BA	**x**	**y**	**z**	**Effect**	**No. Vox.**	**Mass**
Right Middle temporal gyrus	21	54	-4	-7	<0.001	346	1398
Superior Frontal gyrus	10	22	67	-13	0.006	4	12
Medial Frontal gyrus	10	11	70	-7	0.005	5	16
Medial frontal gyrus	10	18	48	4	0.003	11	36
Lingual gyrus	18	22	-67	-2	<0.001	487	1741
Fusiform gyrus	37	40	-37	-13	0.004	5	16
Lentiform nucleus, Putamen	-	14	0	9	<0.001	116	403
Caudate head	-	18	30	4	0.002	11	37
Left Middle temporal gyrus	21	-61	-11	-7	<0.001	325	1177
Medial frontal gyrus	9	0	44	20	<0.001	75	253
Medial frontal gyrus	8	-4	41	42	0.006	5	15

BA = Brodmann area
No Vox. = number of voxels in each cluster
Mass = sum of all the statistical values in the cluster

Localization of voxels with maximum activation from each cluster for the *all sounds > rest* contrast.

Table S2. Neutral Vocalization versus Nonvoice Contrasts, Related to Figure 2

				Infant template Talairach coord.			Activation clusters		
Task / Contrast			**BA**	**x**	**y**	**z**	**Effect**	**No. Vox.**	**Mass**
Neutral vocalization > rest							<0.001		
	Right	Middle temporal gyrus	21	61	0	-7		240	972
		Inferior parietal lobule	40	54	-37	26	0.005	4	13
		Inferior parietal lobule	40	58	-22	26	0.003	6	19
		Medial frontal gyrus	9	7	44	37	0.005	5	16
		Cerebellum, posterior lobe	-	33	-59	-18	0.002	5	16
		Cerebellum	-	7	-56	-2	0.004	7	23
		Cerebellum	-	25	-70	-13	<0.001	10	36
	Left	Middle temporal gyrus	21	-58	-11	-13	<0.001	160	576
		Cingulate gyrus	24	-4	-7	42	0.001	59	192
		Caudate head	-	-7	0	4	<0.001	330	1290
		Cerebellum, anterior lobe	-	-25	-56	-29	<0.001	14	51
Nonvoice > rest									
	Right	Middle temporal gyrus	21	54	-44	9	<0.001	4	18
		Medial frontal gyrus	10	14	56	4	0.005	5	16
		Fusiform gyrus	20	36	-41	-18	0.002	29	94
		Lingual gyrus	18	18	-74	-2	<0.001	54	188
		Lingual gyrus	19	32	-56	4	0.004	15	48
		Caudate head	-	11	26	-2	0.004	8	26
		Caudate	-	29	-33	15	0.003	67	215
		Putamen	-	14	0	-7	0.006	23	72
		Cerebellum, anterior lobe	-	11	-37	-24	0.001	37	128
		Cerebellum, anterior lobe	-	22	-52	-18	0.001	14	48
	Left	Superior temporal gyrus	22	-40	-48	15	<0.001	213	725
		Medial frontal gyrus	10	-7	59	4	0.004	36	115
		Anterior cingulate	32	-18	41	4	0.003	75	240
		Putamen	-	-14	7	-7	0.002	12	40
		Cerebellum, anterior lobe	-	-14	-41	-24	<0.001	109	384
Neutral vocalization > Nonvoice							0.004		
	Right	Superior Temporal Gyrus	38	47	19	-29		6	19
		Middle temporal gyrus (STS)	21	58	7	-18	0.001	19	64
		Medial frontal gyrus	6	14	-4	53	<0.001	36	131
	Left	Medial frontal gyrus	6	-14	-15	53	0.001	19	65
Neutral vocalization < Nonvoice							0.008		
	Left	Superior temporal gyrus	13	-40	-44	15		3	9

BA = Brodmann area
No Vox. = number of voxels in each cluster
Mass = sum of all the statistical values in the cluster

Localization of voxels with maximum activation from each cluster for the *neutral vocalization > rest*, *nonvoice > rest* and *neutral vocalization* versus *nonvoice* contrasts.

Table S3. Correlation with Age, Related to Figure 3

			Infant template Talairach coord.			Correlation clusters			
Task / Contrast		**BA**	**x**	**y**	**z**	**Corr. coefficient**	**p-value**	**No. Vox.**	**Mass**
Neutral vocalization > Nonvoice	Left superior temporal gyrus	22	-51	-44	15	0.632	0.0006	9	5.22

BA = Brodmann area
No Vox. = number of voxels in each cluster
Mass = sum of all the statistical values in the cluster

Table S4. Sad Vocalization > Neutral Vocalization Contrast, Related to Figure 4

				Infant template Talairach coord.			Activation clusters		
Task /Contrast			**BA**	**x**	**y**	**z**	**Effect**	**No. Vox.**	**Mass**
Sad vocalization > neutral vocalization	Left	Gyrus rectus	11	-7	44	-24	0.003	4	13
		Insula	13	-43	4	-2	0.003	3	10

BA = Brodmann area
No Vox. = number of voxels in each cluster
Mass = sum of all the statistical values in the cluster

Localization of voxels with maximum activation from each cluster for the *sad vocalization* versus *neutral vocalizations*.

Supplemental Experimental Procedures

Participants

45 healthy infants (25 girls and 20 boys) aged between 3 and 7 months (age range between 91 and 240 days, mean ± SD = 159 ± 33 days after birth, gestation corrected) were recruited into this study. 3 infants were born at 36 weeks gestation and the remaining infants at full-term. No sedation was used, instead the infants were allowed to fall into natural sleep before entering the MRI scanner. The same infants participated in an fNIRS study using similar stimuli in a separate session within two weeks of the MRI study. Analysis of this data is ongoing and will be presented in a separate paper. All parents gave their written informed consent. The study was approved by the Institute of Psychiatry and South London and Maudsley research ethics committee.

The present study had a global success rate of 47% (21 of the 45 infants who participated in the study), meaning that nearly half the infants who came to the lab were included for analyses. The difficulty in allowing infants to fall asleep in the lab and to remain asleep during the whole session was the main reason for this attrition rate: 24% (11 out of 45) of infants did not fall asleep or woke up during the preparation for the scanner; 16% (7 out of 45) of infants woke up in the scanner before the start of the fMRI session; and for the remaining 13% (6 out of 45), fMRI acquisition was stopped when they woke up. This success rate was achieved by the following measures: sessions scheduled around the infant's nap time, the preparation for the scanner was performed in a dark and quiet room, and steps were taken to lower the likelihood that infants were awoken by the scanner noise (MiniMuffs, sound attenuating foam in the scanner bore, presentation of white noise to mask the start and stop of the scanner, limitation of slew rate). Importantly, all infants with sufficient fMRI data (including two infants with almost complete fMRI data sets of 309 and 280 volumes out of 320) were included in the analyses, and therefore no attrition was based on data itself. The reported data may come from the infants who fell in deeper sleep stages during the testing session and/or those who were less disturbed by sounds while asleep. Auditory activation could have been suppressed to a certain degree in this subgroup of infants, but this did not prevent us from finding stimulus-driven differences in brain activation.

Stimuli

Voice stimuli were chosen from the Montreal Affective Voices [28], the stimuli of the functional localizer of the 'Temporal Voice Areas' available on the Voice Neurocognition Laboratory website (http://vnl.psy.gla.ac.uk/resources_main.php). Some nonvoice stimuli were chosen from the Voice Neurocognition Laboratory website (water sounds), while others were recorded by the authors (toy sounds). Each stimulus sequence lasted 21 seconds and consisted of 7-11 different sounds interleaved by short periods of rest (between 0.47 and 0.75 second). Voices were produced by different adult (male and female) speakers. In each 21 seconds stimulus sequence, the sound volume was gradually increased over a period of 6 seconds to avoid a startle response.

A block design was used to maximize statistical power, in which 21 seconds of auditory stimuli were alternated with 9 seconds of rest. A complete fMRI session comprised 32 blocks (8 in each stimulus category) for a total of 16 minutes. The order of stimulus presentation was weighted so that the neutral vocalization and nonvoice categories appeared more frequently at the beginning of each session. The presentation order aimed to maximize the amount of data in the Voice versus Nonvoice contrast in case an infant could not complete the whole fMRI session.

Testing Procedure

Each family attended an information session during which the testing procedure was demonstrated in a mock scanner. Parents were presented with the stimuli to be used in the study, given an explanation of how infants are prepared for MRI testing and were encouraged to ask questions. A CD of scanner sounds was also given to parents to play at home during their infant's sleep to acclimatize them to these sounds.

On the scanning day, families were invited to come to the laboratory an hour before their infant's nap time to allow the infant to become familiar with the experimental setting and to feed if required. Parents and the infant were then transferred to a quiet room to let the infant fall asleep on a removable MR examination table. When asleep, the infant was swaddled in a cotton sheet, and comfortably positioned in a MedVac Vacuum Immobilization Bag (DFI Medical Solutions), which aimed to reduce movement during the scanning procedure. In order to reduce the noise level perceived by the infant, Natus MiniMuffs Noise Attenuators were placed on the infant's ears and the scanner bore was lined with sound attenuating foam. MR-compatible piezoelectric headphones (http://www.mr-confon.de/en/) were placed on top of the MiniMuffs to reduce residual noise from the MRI scanner and present the stimuli. The sound level of the stimuli was adjusted to a comfortable level for the infants, but that was loud enough to be heard above the residual scanner noise and the MiniMuffs sound attenuation. White noise was presented via the headphones during the scanning procedure (except during the fMRI task) to mask the start and end of the scanner noise.

MRI data were acquired using a GE 1.5 Tesla Twinspeed MRI scanner (General Electric, Milwaukee, WI, USA). 320 T2* weighted gradient echo echo planar multi-slice datasets depicting BOLD (Blood Oxygenation Level Dependent) contrast were acquired in each of 24 noncontiguous near-axial planes (4.0 mm thick with 1.0 mm spacing, 3.5 x 3.5mm in-plane resolution) parallel to the Anterior Commissure-Posterior Commissure (AC-PC) line (TE 57 ms, TR 3000 ms, flip angle 90°, number of signal averages = 1, 16:04 minutes). At the same session a T2 weighted fast spin echo (FSE) dataset was acquired (256x168 rectangular matrix, 2mm slice thickness, 0mm slice gap, field of view=18cm, TR=4500, TE=113ms, echo train length=17). The FSE images were reviewed by a paediatric neuroradiologist to screen for any abnormalities. Gradient rise times were limited in order to reduce the noise of the pulse sequences to approximately 70dB. Daily quality assurance was carried out to ensure high signal to ghost ratio, high signal to noise ratio and excellent temporal stability using an automated quality control procedure [S2]. The body coil was used for RF transmission and an 8-channel head coil for RF reception [S2]. A mcDESPOT sequence [S3] was run in infants who were still asleep and comfortable after the FSE and functional paradigm. Results of this study are presented in a separate paper [S4]. The whole scanning procedure lasted less than 40 minutes and was stopped immediately if the infant awoke and showed discomfort. An experimenter and a parent stood in the scanner room to observe the infant's behaviour at all time and the infant's heart rate was monitored using a pulse oximeter secured on the toe.

Data Analysis

The data was analysed with XBAM software (www.brainmap.co.uk/xbam.htm) using a data-driven approach. Minimization of motion-related artefacts and removal of linear trends was carried out with a rigid-body transform to account for translation and rotation using a spin-history correction [S5]. After registration, the 3D images were realigned at each time point by finding the combination of rotations (around the three axes) and translations (in three

dimensions) that maximized the correlation with an image obtained by averaging the intensity at each voxel over the whole experiment. Then, data were smoothed using a Gaussian filter (7.2 mm isotropic FWHM). Each component of the experimental design was convolved with two gamma variate functions (peak responses at 4 and 8 sec respectively). Then, the best fit between the weighted sum of these convolutions and the time series at each voxel was computed by standard general linear modelling (GLM) and an estimate of the response (beta) obtained for each experimental condition. The infant BOLD response was faster than typical adult responses and was characterized by a significant undershoot after the first peak (Figure S2). Therefore the standard haemodynamic response function (HRF) typically used in the analysis of adult data was not suitable here. To overcome this problem, we adopted a strategy of obtaining HRF information directly from the data, whilst minimizing the statistical bias that could result from this approach. Reasoning that an auditory experiment should produce a dominant auditory response, we obtained the HRF in the auditory cortex for each subject by deconvolution from the mean time-series response in this brain region. For each subject, we then used the mean HRF estimated from all the *other* subjects, thus producing the best estimate of the HRF unbiased by the subject being analysed. The data for each subject were then analysed using standard GLM analysis and the estimated unbiased HRF. This was repeated for all participants.

In the last step of the data analysis, the data were normalized to Talairach space using an infant template as previously described by Dehaene-Lambertz et al [9]. As an infant template was used, the Talairach coordinates provided in the Tables must be treated with caution as they are only an approximation to the adult coordinates.

After normalization, a one sample t statistic was computed at each voxel using the beta estimates for each individual. The significance of this t statistic was then tested at a voxel-wise level by data permutation. Briefly, the signs of the betas were randomly permuted and the resulting permuted betas used to recalculate a one-sample t statistic. Repeating this procedure 40,000 times per voxel produces the distribution of t under the null hypothesis without the requirement for the data to follow a normal distribution – an assumption often violated in small group fMRI data. The significance of the t obtained from the unpermuted data was assessed by reference to the probability distribution of t obtained by data permutation. We used an uncorrected p value of 0.005 for voxel-wise maps.

Methodological Considerations

In this fMRI study we make the first steps in defining the brain regions in young infants activated by human voice processing and the modulatory effect of emotions. We prioritized avoiding type II errors in the statistical analysis over avoiding type I errors (as explained in the data analysis section). For this reason, we set a small cluster size threshold for considering regions of activation to be valid (3 voxels). This, in the context of the whole-brain analysis approach, in which no region of interest was pre-specified, is essential to provide reliable clues for future studies on the development of the brain regions associated with voice processing in young infants. However our approach also increases the risk of missing some potentially important areas.

When we investigated the modulatory effect of emotions, the stimulus presentation order was not completely random (as described in the Methods section). The neutral voice and nonvoice conditions appeared most frequently at the beginning of the study, while the emotion conditions appeared mostly towards the end. This nonrandom presentation order might have influenced the results since habituation to the sounds may have reduced the amplitude of the

activation elicited by the conditions presented later in the protocol [S6]. This could have contributed to the lack of effects found in the emotions versus neutral contrasts. Despite this potential limitation we still found more activation with negative emotions than with neutral vocalizations.

Supplemental References

S1. Richter, W., and Richter, M. (2003). The shape of the fMRI BOLD response in children and adults changes systematically with age. NeuroImage *20*, 1122–1131.

S2. Simmons, A., Moore, E., and Williams, S.C. (1999). Quality control for functional MRI using automated data analysis and Shewhart charting. Magnetic Resonance in Medicine *41*, 1274-1278.

S3. Deoni, S.C., Rutt, B.K., Arun, T., Pierpaoli, C., and Jones, D.K. (2008). Gleaning muticomponent T1 and T2 information from steady-state imaging data. Magnetic Resonance in Medicine *60*, 1372-138.

S4. Deoni, S.C., Mercure, E., Blasi, A., Gasston, D., Thomson, A., Johnson, M., et al. (2010). Mapping infant brain myelination with magnetic resonance imaging. Journal of Neuroscience *31* (2), 784-791.

S5. Bullmore, E.T., Brammer, M.J., Rabe-Hesketh, S., Curtis, V., Morris, R.G., Williams, S.C.R., Sharma, T., and McGuire, P.K. (1999) Methods for diagnosis and treatment of stimulus correlated motion in generic brain activation studies using fMRI. Human Brain Mapping *7*,38-48.

S6. Grill-Spector, K., Henson, R., Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends in Cognitive Sciences *10* (1), 14-23.