Neuropsychologia 122 (2019) 11–19

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Young children's neural processing of their mother's voice: An fMRI study

Pan Liu[a], Pamela M. Cole[a,*,1], Rick O. Gilmore[a], Koraly E. Pérez-Edgar[a], Michelle C. Vigeant[b], Peter Moriarty[b], K. Suzanne Scherf[a]

[a] Department of Psychology, Child Study Center, The Pennsylvania State University, University Park, PA, USA
[b] Graduate Program in Acoustics, The Pennsylvania State University, University Park, PA, USA

ARTICLE INFO

Keywords:
Children
fMRI
Maternal speech
Individual differences

ABSTRACT

In addition to semantic content, human speech carries paralinguistic information that conveys important social cues such as a speaker's identity. For young children, their own mothers' voice is one of the most salient vocal inputs in their daily environment. Indeed, qualities of mothers' voices are shown to contribute to children's social development. Our knowledge of how the mother's voice is processed at the neural level, however, is limited. This study investigated whether the voice of a mother modulates activation in the network of regions activated by the human voice in young children differently than the voice of an unfamiliar mother. We collected fMRI data from 32 typically developing 7- and 8-year-olds as they listened to natural speech produced by their mother and another child's mother. We used emotionally-varied natural speech stimuli to approximate the range of children's day-to-day experience. We individually-defined functional ROIs in children's voice-sensitive neural network and then independently investigated the extent to which activation in these regions is modulated by speaker identity. The bilateral posterior auditory cortex, superior temporal gyrus (STG), and inferior frontal gyrus (IFG) exhibit enhanced activation in response to the voice of one's own mother versus that of an unfamiliar mother. The findings indicate that children process the voice of their own mother uniquely, and pave the way for future studies of how social information processing contributes to the trajectory of child social development.

1. Introduction

The voice is a primary channel for human communication, which has motivated research on the neural basis of voice processing in adults. Unsurprisingly, in adults, voice familiarity modulates voice-sensitive neural networks, with familiar voices eliciting enhanced activation (Belin et al., 2004). In children, one of the most important familiar voices in their lives is that of their mother. A mother's voice is a salient from early in life and it contributes significantly to several aspects of children's development. Indeed, the mother's voice triggers fetal motor and heart rate changes as early as the third trimester (Voegtline et al., 2013), is preferred by newborns in contrast to an unfamiliar female voice (DeCasper and Fifer, 1980; Mehler et al., 1978; Ockleford et al., 1988), and regulates infant cardiac and respiratory activity (Katagiri and Kamikokuryo, 1987; Uchida et al., 2018). Mother's speech enhances young children's attentional and language development (Spinelli et al., 2017), mother's comforts stressed preschoolers (Adams and Passman, 1979) and regulates school-age girls' stress-induced cortisol responses (Seltzer et al., 2012).

In addition to its positive contributions, the mother's voice can also pose potential risk to children. For instance, children from high conflict homes display hypersensitivity to angry stimuli, an effect heightened by their mother's angry voice or face (Graham et al., 2013; Kim and Cicchetti, 2010; Shackman et al., 2007). Shackman et al. (2007) found that physically abused children over-attended to angry facial and vocal cues compared with control children, as indicated by event-related potential (ERP) components. This effect was potentiated when the cues were produced by their abusive mother. Further, the amplitude of the ERP response to abusive mother's angry cues mediated the relation between children's abuse experience and anxiety symptoms. These patterns may also explain vulnerabilities found in children exposed to high levels of interparental conflicts (Cummings and Davies, 2010; Graham et al., 2013). These findings suggest that children's neural processing of vocal cues from significant persons in their lives, such as their mother, may play a mechanistic role that links early experience and later outcomes. To understand the role of familiar voices in children's developmental pathways, it is first necessary to understand the neural basis of children's processing of familiar voices, a subject that

* Corresponding author.
E-mail address: pmc5@psu.edu (P.M. Cole).
[1] This research was supported by a National Institute of Mental Health award (MH104547) to the corresponding author.

https://doi.org/10.1016/j.neuropsychologia.2018.12.003
Received 27 September 2018; Received in revised form 13 November 2018; Accepted 3 December 2018
Available online 05 December 2018

has not been fully investigated in the literature. The goal of the current study is to address this subject by examining whether hearing the natural, emotionally-diverse speech of one's own mother, in contrast to an unfamiliar mother, uniquely modulates the voice-sensitive network in young children.

Our understanding of neural processing of the human voice comes largely from fMRI studies with adults. The identified adult neural network centers on temporal regions, such as the superior temporal gyrus and sulcus (STG, STS) and some fronto-limbic areas (Belin et al., 2004; Frühholz et al., 2016; Schirmer and Kotz, 2006). fMRI studies with infants and children hearing the voices of unfamiliar adults indicate voice-evoked activity in regions similar to that of adults, suggesting that the voice-sensitive neural circuitry emerges very early in life (Blasi et al., 2011; Dehaene-Lambertz et al., 2002; Graham et al., 2013; Raschle et al., 2014; Vannest et al., 2009).

Importantly, a small number of studies indicate that hearing one's mother's voice modulates the voice-sensitive network in infants and children. An fMRI study of sleeping 2-month-olds revealed greater activation in the left posterior temporal lobe, amygdala, and orbitofrontal cortex (OFC) in response to mother's voice of story-reading compared to that of an unfamiliar mother (Dehaene-Lambertz et al., 2010). Greater activation evoked by mothers' versus unfamiliar adult voices was also found for infants using functional Near-Infrared Spectroscopy (NIRS; Imafuku et al., 2014; Naoi et al., 2012). An fMRI study presented 10-year-old children with brief nonsense words produced by their mother and one of two unfamiliar mothers (Abrams et al., 2016). Compared to unfamiliar adult voices, the mother's voice evoked greater activity in the primary auditory cortex, STS, amygdala, insula, nucleus accumbens, OFC, and cingulate. The stimuli used in these studies, however, do not approximate the natural speech that occurs between family members. Two fMRI studies with adolescents more closely approximated that type of experience (Aupperle et al., 2016; Silk et al., 2017). When hearing their mother criticize or praise them, youth with more depressive symptoms showed hyper-activation in the parahippocampus and amygdala in response to maternal criticism, and hypo-activation in those regions in response to maternal praise. However, these more naturalistic studies did not employ the necessary control condition to establish unique patterns associated with mother's voice.

This small but growing body of research on children's neural processing of the mother's voice is limited in several ways. First, most studies lack the control condition of another child's mother as the contrast, thus confounding unfamiliarity and potential factors associated with motherhood. Hormonal changes during pregnancy alter voice muscles, lung capacity and nasal resonance, which influence vocal quality (Cassiraga et al., 2012). In the current study, we asked the mother of each child to serve as an unfamiliar voice for another child, to control for effects of motherhood on the voice. Second, most studies used stimuli that do not approximate the child's real environment, e.g., nonsense syllables spoken in an enjoyable, engaging tone (Abrams et al., 2016) or infant-directed story reading (Dehaene-Lambertz et al., 2010), which do not represent the range of mother's naturally occurring communications. Those communications, more than the voice reading or speaking nonsense syllables, may be important in understanding the mechanistic role of familiar, personally relevant voices in children's development. Thus, we asked mothers to generate emotionally varied natural speech, which mimicked one side of a phone conversation with the spouse, to create stimuli that better approximated children's everyday experiences. Third, many studies of children employ imaging methods that have poor spatial resolution (e.g., NIRS, ERP) because it is hard for young children to remain still. We used the fMRI method, which has greater spatial resolution, and extensively trained children to remain still in the scanner to take full advantage of the high spatial resolution of fMRI.

Our first aim was to delineate the individual variability within the neural network supporting voice processing in young school-age children by using an individual-specific analytical strategy. Substantial individual differences in brains, both anatomical and functional, are underestimated in the traditional, group-level analysis. We aimed to provide evidence of individual variation by identifying individual-specific regions of interest (ROIs) activated by speech stimuli across speaker conditions, using each child's Run 1 data. Second, we aimed to investigate how mother's voice modulates the individual-specific voice-sensitive network, by applying the individual-specific ROIs identified from Run 1 to each individual's data from Run 2 with indices of neural activation extracted for each speaker condition and compared between speakers. To this aim, we focused on a sample of 7- to 8-year-old children, who are old enough to perform fMRI tasks, developing new peer relations but still largely dependent on their caregivers. It also represents a unique developmental window when children can appraise what they hear and overhear. Studying how children of this age process their mother's voice has implications for understanding their relationships and interactions with caregivers, and how these interactions, in turn, influence children's socioemotional development.

Based on the available literature on the neural processing of voices, we predicted that the analyses of Run 1 data would identify a temporal-centered, voice-responsive network as found for adults (Belin et al., 2004; Frühholz et al., 2016; Schirmer and Kotz, 2006) and children (Abrams et al., 2016; Blasi et al., 2011; Dehaene-Lambertz et al., 2010; Graham et al., 2013). Importantly, we further hypothesized that an individual-specific analysis would indicate substantial inter-individual variability in this network, which has not been reported previously. Informed by the neural model of speech prosody processing proposed by Schirmer and Kotz (2006), we focused on the cortical areas of the auditory cortex, superior temporal regions, inferior frontal gyrus (IFG), and orbitofrontal cortex (OFC). Based on our previous work with similar stimuli (Maggi et al.,unpublished data), we also included the amygdala as one of the *a priori* ROIs.

Second, using Run 2 data, we predicted that within the individual-specific regions, one's own mother's voice would elicit enhanced activity relative to an unfamiliar mother's voice. In particular, we expected to observe the enhancement in regions more associated with socio-emotional processing (e.g., STG, the fronto-limbic areas) rather than those involved in primary, acoustic processing (e.g., primary auditory cortex). We believe that such evidence can serve as the foundation for future research investigating children's neural processing as an early step in the developmental pathways by which children's interactions with caregivers and other familiar persons enhance or interfere with their socio-emotional development.

2. Material and methods

This study was part of a larger project investigating children's processing of socio-emotional inputs in their home environment. The data used for the present study were collected from three visits. At visit 1, we recorded mothers' speech to create the stimuli. Visit 2 occurred on average of 26 days later (SD = 18.25), during which fMRI data were collected from children while they listened to speech from their own mother and another mother unfamiliar to them. Visit 3 occurred an average of 27 days (SD = 19.53) after visit 2, during which children listened to the same stimuli and rated the perceived intensity of each stimulus on a 4-point scale (e.g., *how angry/happy/sad does this sound?* 1 = not at all, 4 = very much). These ratings were used as control covariates in the fMRI analyses.

2.1. Participants

Participants were 56 7- and 8-year-old typically-developing children (27 boys; mean age = 8.01 years, SD = 0.50) and their mothers. All children were from two-parent families, with normal hearing, and normal receptive and expressive language abilities as measured by the Clinical Evaluation of Language Fundamentals, 4th edition (Semel et al., 2003): mean score = 110.95, SD = 11.55. All 56 families

Table 1
Scripts used for stimuli recording.

Scenario	Version	
1	A	*Where is the checkbook? It's gone, I can't find it. I don't have it. Do you have it?*
	B	*Do you have the checkbook? You had it last. It's just not here. I'll look for it.*
2	A	*Oh, hi, it's you. When will you be home? Dinner won't be ready then. Okay, I'll fix dinner.*
	B	*I'm fixing dinner. It will take an hour. I have a lot to do. I'll see you later.*
3	A	*Hi, I hoped you'd call. You're running late? I will need some help. Can you change your plans?*
	B	*I could use your help. There's so much to do. Can you change your plans? See you when you get here.*
4	A	*Oh, you're tired? Sorry to hear that. We should talk. About lots of things.*
	B	*Can you talk now? About lots of things. Money, the weekend. Okay, we won't talk now.*

Fig. 1. Abridged illustration of one run of the block-design fMRI task.

completed visit 1, and 53 children (26 boys; mean age = 8.01 years, SD = 0.52) returned for visit 2 for the fMRI task and completed stimuli ratings in visit 3. The fMRI data from 3 children were excluded due to experimenter error, artifacts caused by dental spacers, and aborted scans. Data from an additional 18 children were excluded due to excessive motion in one run or both runs. Eventually, we analyzed fMRI data from 32 children who contributed usable data for both runs.

2.2. Stimuli

At visit 1, we recorded each mother as she spoke scripts adapted from prior research using one-sided phone conversations as stimuli (Cummings, personal communication). Table 1 provides the scripts that involved four typical daily-life scenarios (looking for a checkbook; making dinner; needing help; wanting to talk) with neutral semantic meaning that could be spoken in different tones of voice (angry, happy, sad, and non-emotional). Two versions (A and B) of each scenario with slightly different words were used to avoid repetition. To approximate variations that occur in everyday family communication, mothers produced each script in four prosodies: angry, happy, sad, and non-emotional, which yielded 32 speech stimuli (4 scenarios × 4 prosodies × 2 versions).

A trained graduate student, supervised by an acoustics professor, recorded mothers' speech in a sound isolation room. We used a large diaphragm condenser microphone, sampled at 48 kHz using an M-Audio M-Track USB A/D interface, and Adobe Audition on a laptop computer to record the mother's stimuli. A trained research assistant guided the mothers' practice prior to recording and as needed during recording. Mothers read each script aloud for a minimum of three recordings. The practice helped mothers feel more comfortable with the scripts and voice tones but mothers were allowed to produce the scripts as they felt they would ordinarily. We asked each mother if her productions felt natural to her and we selected recordings that felt natural, had the least noise, and had the greatest prosodic quality for stimulus production. All recorded scripts were edited in Matlab (Mathworks, Inc., Natick, Massachusetts) to render them (1) uniformly 10 s in length by adding (or removing) silent intervals between phrases within each script, and (2) the same loudness by means of uniform A-weighted root mean square (RMS) normalization. The recordings are shared for research and educational uses on Databrary.

2.3. The Speech Listening Task

The Speech Listening Task is a passive listening, block-design, fMRI paradigm that was executed in 2 runs. Each run consisted of 32 task blocks. In each run, for 16 task blocks, the participant's mother was the speaker and in the other 16 task blocks, the mother of another child was the unfamiliar adult speaker. Each speaker read 4 scripts in 4 prosodies (sad, happy, angry, neutral) in each run (to generate 16 blocks). To avoid adaptation effects across the two runs, we used different versions of scripts in the two runs. Each block was 10 s in length. There were 6-second fixation blocks interleaved between task blocks. A unique script was read by the speaker in each block. Each run of the task began with 6 s of fixation (see Fig. 1). Each run ended with a final block of stimuli spoken by the unfamiliar adult in non-emotional prosody to ensure that the task ended on a neutral tone. Data from that last block were excluded from analyses.

Within each run, 4 pseudo-randomized orders of the blocks were created to ensure that (1) the first block heard was a non-emotional, unfamiliar adult voice to eliminate any initial "startle" response; and (2) each type of block (2 speakers × 4 prosodies) followed and preceded every other type of block with equal probability. Each run lasted 538 s (8 m and 58 s). Throughout each run, a circle randomly popped up on the screen around the fixation (approximately twice per block). To monitor task vigilance and wakefulness, children had to press a button upon seeing each circle.

2.4. Procedure

At visit 1, children were introduced to how MRI scanning works, practiced keeping still in a rainbow-colored cloth tube, and visited the imaging center. At visit 2, immediately prior to scanning, children were trained extensively in a mock scanner for approximately 20 min. Each child practiced lying still inside the mock scanner with simulated scanner noise. During this mock session, children were instructed in relaxation breathing, using mental imagery (e.g., lying in bed watching a movie), and provided with feedback about when they moved. Children practiced a version of the fMRI task in the mock scanner in which an unfamiliar female read a storybook and were instructed to press a button when a circle occasionally appeared on the screen. This extensive simulation procedure reduces anxiety and motion during the full scan (Scherf et al., 2013).

After the mock session, children were told they would hear voice recordings of their mother and another child's mother during scanning. The fMRI task was presented and controlled by Matlab using the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Auditory stimuli were delivered through binaural insert earphones of the Sensimetrics Model S14 system (Sensimetrics Corporation, Malden, Massachusetts). In addition, children wore a pair of foam headphones to mitigate scanner noise. The volume of the auditory stimuli was audible and comfortable at a level that was consistent across participants. For each child, the unfamiliar adult voice was selected from the recordings of other mothers in a pseudo-random manner, such that the stimuli from each child's mother were used as the control voice

for another child. To ensure that the acoustic attributes of stimuli were balanced between speakers across all children, we extracted four acoustic parameters that are typically reported in the prosody literature (Banse and Scherer, 1996; Patel et al., 2011): mean fundamental frequency (f0), SD of f0, harmonics-to-noise ratio, and speech rate, and compared them between the two speaker conditions. As the mother vs. unfamiliar adult comparison of the fMRI data was conducted in Run 2 data only, the acoustic analysis was also focused on the stimuli of Run 2 in the same participants that contributed usable imaging data (N = 31). Multivariate analysis with *speaker* as the independent factor indicated no speaker effect for the four parameters together, $F(4, 57) = 0.22$, $p = .92$, or independently: mean f0, $F(1,60) = 0.03$, $p = .85$; SD of f0, $F(1,60) = 0.04$, $p = .84$; harmonics-to-noise ratio, $F(1,60) = 0.18$, $p = .68$; speech rate, $F(1,60) = 0.46$, $p = .50$.

2.5. fMRI data acquisition, processing, and analyses

Participants were scanned using a 3-T Siemens Prismafit scanner with a 12-channel head coil (Siemens Medical Solutions, Erlangen, Germany). High-resolution T1-weighted structural images were first collected using a magnetization prepared gradient echo sequence (MPRAGE) with the following parameters: 192 1 mm slices, TR = 2300 ms, TE = 2.28 ms, flip angle = 20°, FoV = 256 mm, voxel size = 1 × 1 × 1 mm, T1 = 900 ms. The functional data were collected using echo-planar imaging (EPI) by Generalized Autocalibrating Partial Parallel Acquisition (GRAPPA) and included 40 slices that were aligned approximately 30° perpendicular to the hippocampus, which is effective for maximizing signal-to-noise ratios in the medial temporal lobes (Whalen et al., 2008), where some of our ROIs are located. The scan parameters were as follows: TR = 2000 ms, TE = 25 ms, flip angle = 80°, FoV = 210 mm, voxel size = 3 × 3 × 3 mm, iPAT = 2.

Raw fMRI data were preprocessed using BrainVoyager QX v2.3 (Brain Innovation, Masstricht, The Netherlands). For each functional run, the first 3 volumes were removed to discard unsteady signals. The remaining 266 functional volumes were 3D motion corrected, spatially smoothed (4 mm), and high-pass filtered to remove low-frequency signals (General Linear Model Fourier basis 5 cycles). Children who exhibited spikes in motion > 3 mm on any TR on any vector of either run were excluded from subsequent analyses (N = 18). Each participant's functional images were co-registered to his/her structural images in the native space, and then normalized to the Talairach space. Thirty-two children generated usable data for both runs of the task (16 boys; mean age = 8.07, SD = 0.52).

The central question of this work is whether the speaker identity, and one's own mother's voice in particular, modulates activity in the network of regions activated by the human speech. To address this question, we borrowed an analytic approach that is often used in vision science to investigate the representational properties of regions in the ventral visual pathway (Berman et al., 2010; Fox et al., 2009; Saxe et al., 2006; Scherf et al., 2011) and that represents best practices in analytic approaches to working with fMRI data in datasets with small to moderate-sized samples (see Poldrack et al., 2017). Specifically, we used the data acquired in Run 1 to independently localize the network of regions activated by the human voice for each participant. Then within each of the regions identified in each participant, we used the data acquired in Run 2 to investigate the extent of modulation of activation as a function of speaker identity (i.e., the effect size of the speaker – mother, unfamiliar adult). This analytic approach is critical for ensuring that the voxel selection process for defining regions associated with human speech processing is *independent* from process of estimating the effect size of the speaker identity (mother, unfamiliar adult) on activation within those regions. When these processes are not independent, the modulatory effect size can be inflated because of double-dipping (see Poldrack and Mumford, 2009). A similar localizer approach is used in language processing studies (Fedorenko et al., 2010), but the stimuli and contrasts for defining ROIs in those studies focus on word-reading and sentence-level processes, and not on auditory processing of human speech. No prior study of speaker identity effects has used this conservative approach to safeguard against double-dipping and inflated effect sizes.

2.5.1. Individual-level whole-brain GLMs on data of Run 1

We first conducted individual-level fixed-effects whole-brain voxel-wise General Linear Model (GLM) on the time series data from Run 1 for each of the 32 children who generated usable data for both runs, with the speaker as the fixed regressor. The regressor was convolved by a canonical hemodynamic response function. We identified human voice related regions using the weighted contrast *All speech (mother + control) > Baseline (fixation intervals)*. The functional ROIs to be identified in each individual participant were informed by Schirmer and Kotz's (2006) neural network model of speech processing, our recent fMRI study with children using highly similar speech stimuli (Maggi et al.,unpublished dataa), as well as previous findings in adults (Belin et al., 2004; Frühholz et al., 2016) and children (Abrams et al., 2016; Blasi et al., 2011; Dehaene-Lambertz et al., 2010; Graham et al., 2013). Our *a priori* ROIs included anterior and posterior auditory cortex (AC), superior temporal gyrus (STG), OFC, IFG, and amygdala. Other regions that were reported in previous work but were not activated in our group-level activation map (e.g., the insula) were not included. These ROIs were defined separately in each participant in each hemisphere and were corrected for false positive activation at the whole-brain level False Discovery Rate (FDR)-corrected $q < 0.05$ (Genovese et al., 2002; Eklund et al., 2016).

Each functionally-defined ROI was defined as the set of contiguously activated voxels (i.e., neighboring voxels sharing a face or edge) that satisfied the following anatomical boundaries. The anterior AC ROI was defined as the set of contiguous voxels within the Heschl's gyrus, with Heschl's sulcus as the anterior border and the first transverse sulcus as the posterior border (See Buchanan et al., 2000). The posterior AC ROI was defined as the cluster of voxels immediately posterior to the anterior AC region, with the first transverse sulcus as the anterior border and the bifurcation of the Sylvian fissure as the posterior border. The STG ROI was defined as the set of activated voxels between the horizontal posterior segment of the superior temporal sulcus and the lateral fissure, but did not extend into the ascending posterior segment of STG. The anterior boundary of the STG region was where the ascending segment of the intraparietal sulcus intersected the lateral fissure.

The IFG ROI was defined as the cluster of voxels within the lateral inferior portion of the prefrontal gyrus, with inferior frontal sulcus as the superior border, lateral fissure the inferior border, and inferior precentral sulcus the posterior border. The OFC ROI was defined as the set of activated voxels in the lowermost portion of prefrontal cortex immediately above the orbits. For both prefrontal ROIs, similar locations have been reported in previous studies of speech processing (e.g., Dehaene-Lambertz et al., 2010; Sander et al., 2005; Wildgruber et al., 2006). Finally, the amygdala was defined as the cluster of speech-sensitive voxels within the grey matter structure. Any active voxels that extended into to the surrounding areas including the hippocampus or lateral ventricle were excluded.

In the current data, these ROIs were defined for each individual with his/her functional images overlaid on his/her own structural images, to fully account for the individual variability in brain anatomy. Each ROI was quantified in terms of the total number of significantly active voxels (size), the location of the centroid of the ROI (in Talairach coordinates), and magnitude of activation (beta weights).

Table 2
Individually-defined functional ROIs: the number of participants (N) that showed significant activation for each ROI; the mean (SD) values of Talairach coordinates and number of voxels for each ROI across participants.

		N	Mean(SD) coordinates of centroid			Mean(SD) # of voxels
			x	y	z	
Anterior Auditory Cortex	*Left*	31	− 51(5)	− 24(3)	11(1)	3584(2221)
	Right	30	53(4)	− 22(4)	11(1)	3527(2077)
Posterior Auditory Cortex	*Left*	31	−53(5)	−40(2)	10(3)	1996(1769)
	Right	23	56(4)	−40(2)	10(4)	1577(1350)
STG	*Left*	31	−54(3)	−26(1)	3(1)	3305(1500)
	Right	30	54(3)	−25(3)	2(1)	3506(1577)
IFG	*Left*	24	−40(5)	29(5)	0(4)	802(888)
	Right	23	40(9)	29(4)	−1(4)	849(927)
OFC	*Left*	13	−29(4)	20(4)	−13(4)	353(365)
	Right	14	28(9)	20(7)	−13(5)	350(385)
Amygdala	*Left*	6	−22(6)	−4(5)	−14(3)	188(262)
	Right	6	20(3)	−3(4)	−15(3)	184(150)

2.5.2. Evaluating effects of speaker identity within the voice-sensitive network on data of Run 2

To evaluate the potential modulatory effects of speaker identity on the activation profile within the speech-sensitive ROIs, we computed individual-level ROI-based GLMs with the fixed factor of *speaker*, in each of the individually defined ROIs in each participant separately. To control for potential differences in the perceived emotional intensity between the mother's and unfamiliar adult's voices that might influence neural responses, the intensity rating of each stimulus from Run 2 was included as a vector of variable of no interest in each GLM. Of note, for the non-emotional blocks, children also rated "*how angry/happy/sad does this sound*"; the three rating scores were averaged for each non-emotional block to indicate the subjective rating of the non-emotional stimuli. From each of these independently defined ROIs of each participant, we extracted the mean beta weights for each of the two speaker conditions across both emotional and neutral blocks. Finally, to examine the effects of speaker identity across the network, we submitted the beta weights from pairs of bilateral ROIs to separate linear mixed models with the fixed effects of speaker (mother, unfamiliar adult) and hemisphere (left, right), and the random factor of participant.

3. Results

3.1. Individual-specific activation map using data from Run 1

Results from the individual-level whole-brain GLMs of 32 children are shown in Table 2. In individual-level GLMs, no participant showed significant activation in all *a priori* ROIs, rendering the sample size different across ROIs.

3.2. Speaker identity effect

One participant was excluded from the speaker identity analyses because the intensity rating data were missing due to an experimenter error, rendering an eventual sample size of 31 for this analysis. The results of six linear mixed-effects regression models for each of the six bilateral ROIs are plotted in Fig. 2. The main effect of *speaker identity* was significant or marginally significant in bilateral posterior AC, $F(1, 382.87) = 4.33, p = .038, r^2 = 0.11$, bilateral IFG, $F(1, 338.84) = 4.09$, $p = .044$, $r = 0.11$, and bilateral STG, $F(1, 439.05) = 3.45$, $p = .064$, $r = 0.09$. As shown in Fig. 2, mother's voice elicited greater activation than the unfamiliar voice in all these ROIs. Speaker identity did not modulate activation in anterior AC, OFC, or the amygdala (p's > 0.31). To correct for multiple comparisons, we applied the Benjamini-Hochberg procedure with an FDR of 0.15 to the six models, after which the *speaker identity* effect in the three bilateral ROIs (posterior AC, STG, and IFG) remained significant.[3] No significant *speaker* × *hemisphere* interaction was observed (p's > 0.66).

4. Discussion

Neural processing of the human voice appears to begin prenatally and has important influences throughout childhood. Using natural speech samples that were more representative of children's everyday experiences than stimuli typically used in affective prosody processing research, we provide the first evidence that the voice-sensitive network of 7- and 8-year-old children is modulated differently by the familiar voice of their mother relative to the voice of an unfamiliar mother. As predicted, we observed greater neural activation when the voice was that of the child's mother in a network that meta-analyses define as the speech processing network in adults. There is ample evidence that young school age children are distressed by inter-parental anger (Cummings and Davies, 2010), and evidence that their attributions about inter-parental conflicts mediate the effects of conflict exposure on their adjustment (Grych et al., 2000; Shackman et al., 2007). Our evidence identifies young school age children's neural sensitivity to the voice of a parent, in our case the mother; it may then be possible to assess individual differences in how children's voice- and speech-sensitive networks are modulated differently in the context of different family environments or in terms of young children's interpretations of parental conflicts they overhear or witness. More broadly, our findings offer a foundation for future work exploring the developmental pathways by which children's interactions with caregivers enhance or interfere with their neural processing of socio-emotional information and the consequences of those effects on children's development.

First, our individual-level whole-brain GLM yielded distributed significant activation for the *All speech > Baseline contrast*. Speech-elicited activation across mothers' and control voices was observed in both the auditory areas (anterior and posterior auditory cortex, STG) and the fronto-limbic regions (IFG, OFC, amygdala). This pattern of activation converges with the modest literature in children's processing of human voices (Abrams et al., 2016; Blasi et al., 2011; Dehaene-

[2] Effect size, $r = \sqrt{F/(F + df_{error})}$ (Rosnow and Rosenthal, 2003).

[3] The six p values from the six models were ordered from smallest to largest. For each p value, a Benjamini-Hochberg critical value was calculated as $(i/m) \times q$, i = the rank of the p value, $m = 6$ (total number of models), $q = 0.15$ (FDR). The largest p value that is smaller than the corresponding Benjamini-Hochberg critical value is considered significant; all other p values that are smaller than this largest p value are also considered significant (Benjamini and Hochberg, 1995; McDonald, 2014).

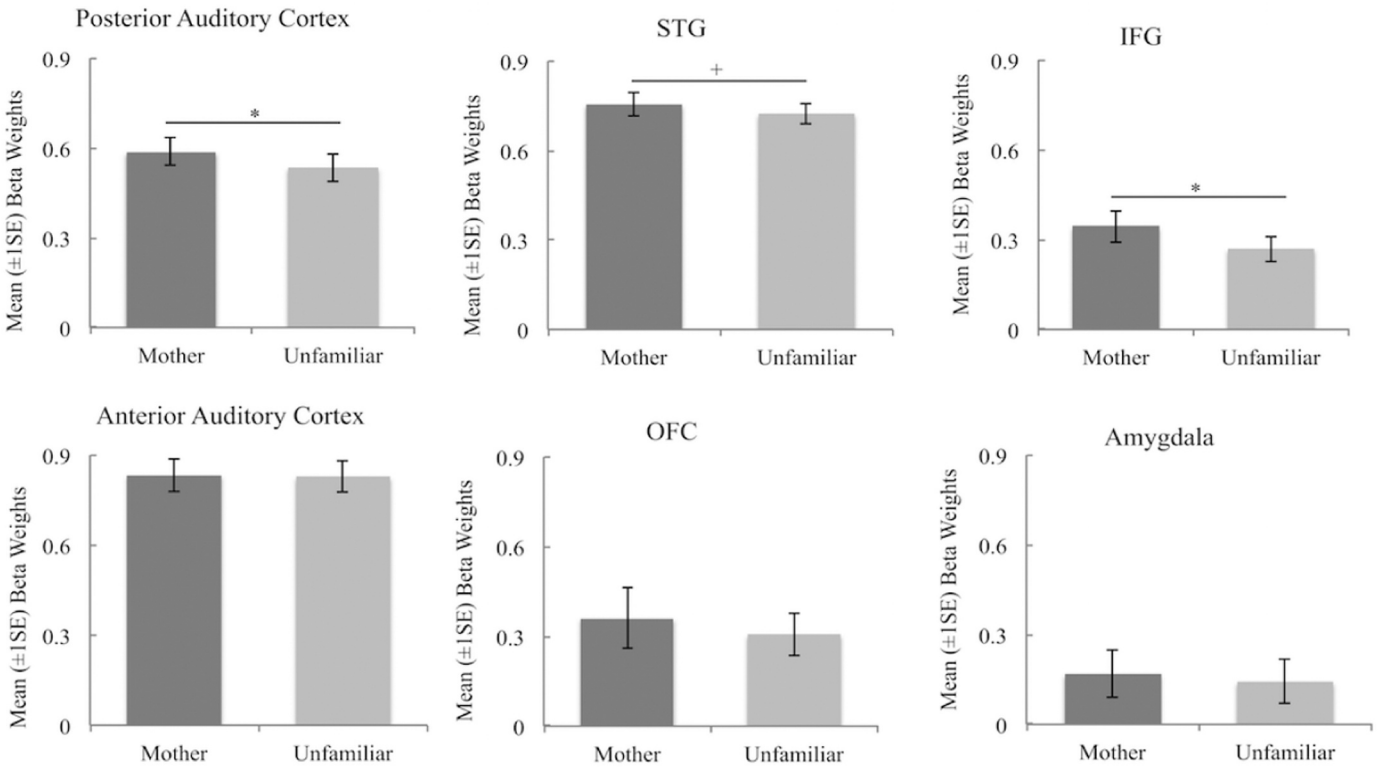

Fig. 2. Beta weights of Run 2 data extracted from individual-specific ROIs defined by Run 1 for the two speaker conditions across hemispheres (*, $p < .05$, $^{+}$, $p < .1$).

Lambertz et al., 2010; Graham et al., 2013; Maggi et al., unpublished data), as well as previous findings in adults (Belin et al., 2004; Frühholz et al., 2016; Schirmer and Kotz, 2006).

Our observed patterns of neural activation converge with a proposed neural network model for speech prosody processing in adults (Schirmer and Kotz, 2006). This model suggests that speech processing occurs in three stages. First, there is audio-sensory processing of acoustic properties via the auditory cortex. Next, the model suggests an auditory "what" pathway in superior temporal areas that integrates socially salient acoustic cues to derive socio-emotional meaning. Third, that socially-salient information is processed in regions associated with higher-order cognitive processing, namely in the IFG and OFC areas; at this level socially-salient information is processed linguistically, and evaluative judgments are made. Further, these prefrontal regions are connected extensively with other cortical and subcortical areas; they project information from multiple sensory modalities and play significant regulatory roles in top-down control processes (Corbetta, 1998; Corbetta and Shulman, 2002; Fox and Pine, 2012; Schirmer and Kotz, 2006).

In addition to cortical regions, we also observed a subcortical area, the amygdala, activated in a subset of children. Although the Schirmer and Kotz (2006) model focuses on cortical processing, activation of subcortical regions, especially limbic areas, have been reported for vocal stimuli with both adults (Bach et al., 2008; Morris et al., 1999; Leitman et al., 2010; Mitchell et al., 2003; Sander et al., 2005) and children (Abrams et al., 2016; Dehaene-Lambertz et al., 2010; Maggi et al., unpublished data). Limbic areas are known to play critical roles in processing socially and emotionally salient information (Adolphs, 2009). The present findings suggest that the neural correlates of socially-salient speech processing in at least some young school age children are similar to those reported for adults, which provides a basis for future comparative studies investigating the neural processing of children of different ages and adults.

In addition to the convergence with published evidence, our individualized analytical approach provides novel evidence that substantial individual variability exists within children's neural network underlying the processing of socially salient speech stimuli. In particular, of the 32 children, most (N = 31 or 30) showed significant activation in the auditory cortex and STG, except for right posterior AC, for which 23 children displayed activation). For the prefrontal areas, IFG (N = 24 or 23) and OFC (N = 13 or 14), the number of children showing significant activation decreased, and only 6 children displayed significant activation in bilateral amygdala.

The finding that only a few child participants met criteria for identifying activation in the frontolimbic ROIs is not surprising. To avoid inducing motion, the task only required passive listening to auditory stimuli, without an explicit judgment or response. As such, the task may have evoked activation in sensory processing regions but with relatively weaker activation in regions that involve emotional salience. Previous studies involving passive processing of vocal or facial stimuli, including our own (Maggi et al., unpublished data), failed to observe significant amygdala activation at the group versus individual level (Fu et al., 2017; Monk et al., 2006). Another possible explanation for how few children showed significant activation in the amygdala may be the fact that it is technically difficult to obtain good signal-to-noise ratio of small neural structures due to factors such as signal dropout incurred by magnetic inhomogeneity and influences from nearby vessels and air-filled cavities (Boubela et al., 2015). For children, there may be more heterogeneity due to these factors. is needed to understand whether task or other imaging factors are at play, or whether the amygdala is only activated for some children as they process familiar voices of family members. Variations such as child-directed versus overheard interparental speech may also modulate or amplify amygdala activity. Yet another possibility is that the observed amygdalar activation reflects the relative immaturity of children's neural systems, such as incomplete myelination of axons in these regions (Paus, 2005). While no findings from children have been reported, amygdala activation elicited by emotional stimuli in adults was positively correlated with white matter integrity in the amygdala-prefrontal pathway, which was in turn modulated by the degree of myelination (Kim and Whalen, 2009).

Future research that combines functional and anatomical imaging data in children will help disentangle the relations between anatomical maturity and functional neural activities in the developing brain. Moreover, future research with larger samples of children can address the extent to which task factors, technical factors, neural developmental factors, or environmental variations are at play. Nonetheless, although individual variation is largely understudied in the imaging literature, our findings are the first to directly map the variant topography of speech-elicited activation among children using an individual-specific analytical approach.

Based upon the individualized functional ROIs defined by Run 1 data, we compared children's activation levels between the two speaker conditions with Run 2 data. As hypothesized, children displayed significantly (or marginally significant) higher activation in response to their own mother's voice versus an unfamiliar mother's voice in bilateral posterior AC, STG, and IFG ROIs. Similar yet non-significant patterns between speakers were observed in bilateral OFC and amygdala. For these two ROIs, it may be that the analyses were underpowered due to small sample sizes (N = 14 and 13 for bilateral OFC, N = 6 and 6 for bilateral amygdala). Overall, these observations of speaker differences showed compatibility with findings of recent developmental fMRI work comparing mothers' versus control voices: multiple brain regions are preferentially activated by one's mother's voice as opposed to the control voice, in both the auditory and frontolimbic areas (Abrams et al., 2016; Dehaene-Lambertz et al., 2010; Imafuku et al., 2014; Naoi et al., 2012). These findings are also consistent with findings from adults that social cues from personally relevant sources elicit greater activation across multiple brain regions (e.g., Nakamura et al., 2001; Natu and O'Toole, 2011; Shah et al., 2001).

Interestingly, however, no speaker difference, not even at trend-level, was observed for bilateral anterior AC ROIs, even though data for these regions were well powered (N = 29 and 30). Anterior AC plays a role in the initial elementary processing of the acoustical properties of the vocal stimuli (Zatorre and Belin, 2001). As proposed in Schirmer and Kotz's model (2006), this area is activated during the very first stage of speech processing and is probably not sensitive to socially relevant information such as speaker identity. It is only when the analyzed acoustic cues enter areas within the "what" pathway, such as the posterior AC and STG, that the socially relevant features begin to be extracted and identified (i.e., stage two of the Schirmer and Kotz's model). Indeed, our data showed that these brain areas begin to differentiate between the two speakers. The literature suggested that the posterior AC, in comparison with the anterior AC, is more tuned to acoustic features that may signify context-relevant salience (Jääskeläinen et al., 2004). The STG is known as a critical region that integrates salient acoustic parameters to form socio-emotional representations (Frühholz et al., 2011), and sub-serves an anterior-wise information flow that relays to higher-order, prefrontal regions (Schirmer and Kotz, 2006).

Finally, socially salient information feeds into higher-order cognitive processing in the prefrontal areas for further integration and evaluation (Ethofer et al., 2006). The individual's discrimination of the speaker identity and personal relevance is maintained during this stage. Specifically, the IFG area plays a role in higher-order evaluation of the prosodic information and response selection (Leitman et al., 2010). The OFC, on the other hand, has a more direct, reciprocal relation with the amygdala and modulates limbic responses to maintain goal-directed behavior (Corbetta and Shulman, 2002; Fox and Pine, 2012). As discussed earlier, the weaker activation and absence of significant speaker-linked differences in OFC and amygdala may be partly due to the nature of the passive listening task that we used. Performing a more explicit task, such as making judgments of the speaker identity or the emotion of the stimuli, might evoke more extensive or greater activation (Vannest et al., 2009). Further, our stimuli imitated overheard speech directed to a third person rather than speech directed to the child. This type of speech stimuli may contribute to weaker activation in frontolimbic regions. Speech directed toward the child, especially from the child's mother, may have even greater personal significance. Future studies with larger samples that can manipulate the target of speech are needed to examine these possibilities.

As proposed by Schirmer and Kotz's model (2006), situational or individual significance, including personal relevance, could enhance or facilitate multiple sub-processes of speech processing mediated by multiple brain structures. It may involve both bottom-up, stimulus-driven mechanisms and top-down, attentional control mechanisms. Studies on sensory processing suggested that when a certain type of sensory stimuli became salient (e.g., through behavioral learning), the neural representation of those stimuli in the sensory cortex would be strengthened, presumably to facilitate goal-directed behaviors (Karni et al., 1998; Wang et al., 1995). In a similar vein, for young children, the personally relevant features of their mother's voice may directly enhance its representation and processing throughout children's speech-sensitive circuit. Further, as suggested by Abrams et al. (2016), hearing one's mother's voice, even when not directed toward themselves, might constitute a rewarding stimulus for young children. This nature of a mother's voice might also contribute to the enhanced activation observed in the prefrontal and limbic areas, which are implicated in the reward circuits of the human brain. However, their study used mother's speech spoken in an engaging and enjoyable tone, which might sound more "rewarding" or comforting to young children than our stimuli produced in various, including negative, prosodies.

It is worth reiterating that our study design used the mother of another child for the control condition. This approach took into account that mothers' voices may differ from that of women who have not been mothers, and enhanced our ability to attribute observed neural activity to the nature of one's own mother. Moreover, the use of emotionally-diverse, natural speech stimuli provided a closer approximation of children's everyday experiences. As such, our study provides a novel methodological template, including the well-controlled stimuli, design, and the individualized analytical approach, for investigating social information processing in children.

In terms of limitations, our sample size was relatively small. In particular, the individual-specific analytical approach rendered significantly smaller sample sizes for certain ROIs (e.g., OFC, amygdala) and prevented effective statistical evaluation for these regions. Future studies will greatly benefit from larger sample sizes, especially as the current literature is moving toward an individual-specific approach as opposed to the traditional, group-averaged analysis (Poldrack, 2017). Future study design that enables Multi-Voxel Pattern Analysis will also provide novel information for the neural substrates of maternal speech processing. Further, in contrasting a child's own mother's voice with an unfamiliar mother's voice, we cannot isolate the "uniqueness" of the mother's voice from its familiarity to children, and cannot determine to what extent the observed effect was driven by the maternal uniqueness or familiarity. More work has been done on this issue in the face processing literature (e.g., Dai et al., 2014; Gobbini et al., 2004). For example, a study examined face processing in children who were raised by their biological mothers and maternal aunts and were therefore equally familiar with their mother's and aunt's faces. Regardless of the equal familiarity, enhanced face-evoked ERP activity was found in response to the mother's versus the aunt's faces, which may reflect the uniqueness of the facial cues of one's biological mother (Dai et al., 2014). Based on these observations in the facial modality, we speculate that similar maternal uniqueness may also exist in the vocal domain. Future studies that compare the mother's voice with the voice of a familiar celebrity without personal relevance, or with the voice of another familiar caregiver (e.g., the father), might help unpack the effects of uniqueness and familiarity.

Finally, the personal relevance of mother's voice might also interact with the emotional valence of the speech. For instance, children might be particularly sensitive and alerted when they hear their mother gets

angry. Shackman et al. (2007) found that physically abused children show an attention bias to angry cues but not happy or sad cues, and this effect was particularly amplified when the angry cue was from their abusive mother. Further, children's heightened attention to mother's angry cues mediated the association between the abuse experience and emergence of anxiety symptoms. In the context of emotional development, understanding the neural processing of maternal angry (and other negative) vocal stimuli is critical to address the potential neurocognitive mechanisms that mediate the pathway from children's early exposure to emotional inputs at home to their later outcomes of socio-emotional functions. In particular, using speech stimuli directed to a third person, especially involving angry and sad tones, has implications for studying the influence of inter-parental conflict on children's emotion processing and emotional development (Cummings and Davies, 2010).

In conclusion, with a carefully manipulated design and ecologically valid speech stimuli that have not been used in the previous literature, this study investigated the neural network underlying 7- and 8-year-old children's representation and processing of their own mother's voices. By taking an individualized, conservative analytic strategy, we for the first time delineated individual differences within the speech-sensitive network, and observed enhanced activity in response to speech from one's own mother versus an unfamiliar mother. Our findings contributed significant novel evidence to the knowledge of how the developing brain represents and processes human speech in general, and their mother's speech in particular. Findings of the current study lay the foundation for future studies that explore how children's interactions with their caregivers, and the emotional inputs from their immediate environment in general, enhance or impede their neurocognitive processing of socio-emotional information, which may in turn shape the developmental trajectory of their socio-emotional functions.

Conflict of interest and acknowledgements

This work was supported by a grant from the National Institute of Mental Health (MH104547) to Dr. Pamela M. Cole. The funding source had no involvement in study design, data collection and analysis, and preparation and submission of the manuscript. The authors report no relevant financial interests or conflicts of interest. The authors would like to thank the Social, Life, & Engineering Sciences Imaging Center and the Child Study Center at The Pennsylvania State University for their support of research facilities and resources. We also want to thank Rachel Wolf and other research staff who were dedicated to data collection. Special thanks are extended to the families who participated in our study.

References

Abrams, D.A., Chen, T., Odriozola, P., Cheng, K.M., Baker, A.E., Padmanabhan, A., Ryali, S., Kochalka, J., Feinstein, C., Menon, V., 2016. Neural circuits underlying mother's voice perception predict social communication abilities in children. Proc. Natl. Acad. Sci. USA 113 (22), 6295–6300.

Adams, R.E., Passman, R.H., 1979. Effects of visual and auditory aspects of mothers and strangers on the play and exploration of children. Dev., Psychol. 15 (3), 269–274.

Adolphs, R., 2009. The social brain: neural basis of social knowledge. Annu Rev. Clin., Psychol. 60, 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514.

Aupperle, R.L., Morris, A.S., Silk, J.S., Criss, M.M., Judah, M.R., Eagleton, S.G., Kirlic, N., Byrd-Craven, J., Phillips, R., Alvarez, R.P., 2016. Neural responses to maternal praise and criticism: relationship to depression and anxiety symptoms in high-risk adolescent girls. NeuroImage: Clin. 11, 548–554.

Bach, D.R., Schachinger, H., Neuhoff, J.G., Esposito, F., Di Salle, F., Lehmann, C., Herdener, M., Scheffler, K., Seifritz, E., 2008. Rising sound intensity: an intrinsic warning cue activating the amygdala. Cereb. Cortex 18, 145–150.

Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. J. Pers. Soc. Psychol. 70, 614–636.

Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B. Stat. Methodol. 57, 289–300.

Berman, M.G., Park, J., Gonzalez, R., Polk, T.A., Gehrke, A., Knaffla, S., Jonides, J., 2010. Evaluating functional localizers: the case of the FFA. NeuroImage 50 (1), 56–71.

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vision. 10, 433–436.

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G.J., Renvali, V., Deoni, S., Gasston, D., Williams, S.C.R., Johnson, M.H., Simmons, A., Murphy, D.G., 2011. Early specialization for voice and emotion processing in the infant brain. Curr. Biol. 21 (14), 1220–1224.

Boubela, R.N., Kalcher, K., Huf, W., Seidel, E.M., Derntl, B., Pezawas, L., Moser, E., 2015. fMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. Sci. Rep. 5, 10499.

Buchanan, T.W., Lutz, K., Mirzazade, S., Specht, K., Shah, N.J., Zilles, K., Jäncke, L., 2000. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. Cogn. Brain Res. 9 (3), 227–238.

Cassiraga, V.L., Castellano, A.V., Abasolo, J., Abin, E.N., Izbizky, G.H., 2012. Pregnancy and voice: changes during the third trimester. J. Voice 26 (5), 584–586.

Corbetta, M., 1998. Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? Proc. Natl. Acad. Sci. USA 95, 831–838.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3 (3), 201–215.

Cummings, E.M., Davies, P.T., 2010. Marital Conflict and Children: An Emotional Security Perspective. Guilford, New York.

Dai, J., Zhai, H., Wu, H., Yang, S., Cacioppo, J.T., Cacioppo, S., Luo, Y.J., 2014. Maternal face processing in Mosuo preschool children. Biol. Psychol. 99, 69–76.

DeCasper, A.J., Fifer, W.P., 1980. Of human bonding: newborns prefer their mothers' voices. Science 208 (4448), 1174–1176.

Dehaene-Lambertz, G., Dehaene, S., Hertz-Pannier, L., 2002. Functional neuroimaging of speech perception in infants. Science 298 (5600), 2013–2015.

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., Dehaene, S., 2010. Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. Brain Lang. 114 (2), 53–65.

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. USA 113 (28), 7900–7905. https://doi.org/10.1073/pnas.1602413113.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., Wildgruber, D., 2006. Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. Neuroimage 30, 580–587.

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. J. Neurophysiol. 104 (2), 1177–1194. https://doi.org/10.1152/jn.00032.2010.

Fox, C.J., Iaria, G., Barton, J.J., 2009. Defining the face processing network: optimization of the functional localizer in fMRI. Hum. Brain Mapp. 30 (5), 1637–1651.

Fox, N.A., Pine, D.S., 2012. Temperament and the emergence of anxiety disorders. J. Am. Acad. Child Adolesc. Psychiatry 51 (2), 125–128.

Frühholz, S., Ceravolo, L., Grandjean, D., 2011. Specific brain networks during explicit and implicit decoding of emotional prosody. Cereb. Cortex 22 (5), 1107–1117.

Frühholz, S., Trost, W., Kotz, S.A., 2016. The sound of emotions-towards a unifying neural network perspective of affective sound processing. Neurosci. Biobehav. Rev. 68, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002.

Fu, X., Taber-Thomas, B., Pérez-Edgar, K., 2017. Frontolimbic functioning during threat-related attention: relations to early behavioral inhibition and anxiety in children. Biol. Psychol. 122, 98–109.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15, 870–878.

Gobbini, M.I., Leibenluft, E., Santiago, N., Haxby, J.V., 2004. Social and emotional attachment in the neural representation of faces. Neuroimage 22 (4), 1628–1635.

Graham, A.M., Fisher, P.A., Pfeifer, J.H., 2013. What sleeping babies hear: a functional MRI study of interparental conflict and infants' emotion processing. Psychol. Sci. 24, 782–789.

Grych, J.H., Fincham, F.D., Jouriles, E.N., McDonald, R., 2000. Interparental conflict and child adjustment: testing the mediational role of appraisals in the cognitive-contextual framework. Child Dev. 71 (6), 1648–1661.

Imafuku, M., Hakuno, Y., Uchida-Ota, M., Yamamoto, J., Minagawa, Y., 2014. "Mom called me!" Behavioral and prefrontal responses of infants to self-names spoken by their mothers. Neuroimage 103, 476–484.

Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levänen, S., Belliveau, J.W., 2004. Human posterior auditory cortex gates novel sounds to consciousness. Proc. Natl. Acad. Sci. USA 101 (17), 6809–6814. https://doi.org/10.1073/pnas.0303760101.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M.M., Turner, R., Ungerleider, L.G., 1998. The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. Proc. Natl. Acad. Sci. USA 95 (3), 861–868.

Katagiri, K., Kamikokuryo, S., 1987. The development of heart rate responses to environmental auditory stimuli in infants during the first six months of life. Psychiatr. Neurol. Paediatr. Jpn. 27, 227–236.

Kim, M.J., Whalen, P.J., 2009. The structural integrity of an amygdala-prefrontal pathway predicts trait anxiety. J. Neurosci. 29 (37), 11614–11618. https://doi.org/10.1523/JNEUROSCI.2335-09.2009.

Kim, J., Cicchetti, D., 2010. Longitudinal pathways linking child maltreatment, emotion regulation, peer relations, and psychopathology. J. Child Psychol. Psychiatry 51 (6), 706–716. https://doi.org/10.1111/j.1469-7610.2009.02202.x.

Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox-3? Percept. 36 ECVP Abstract Supplement.

Leitman, D.I., Wolf, D.H., Ragland, J.D., Laukka, P., Loughead, J., Valdez, J.N., Javitt, D.C., Turetsky, B.I., Gur, R., 2010. "It's not what you say, but how you say it": a reciprocal temporo-frontal network for affective prosody. Front. Hum. Neurosci.

4, 19.
Maggi, M.C., Cole, P.M., Elbich, D., Gilmore, R.O., Pérez-Edgar, K., Scherf, K.S., unpublished data. Hearing emotions: School-aged children's neural processing of the human voice and affective prosody.
McDonald, J.H., 2014. Handbook of Biological Statistics, third ed. Sparky House Publishing, Maryland.
Mehler, J., Bertoncini, J., Barriere, M., Jassik-Gerschenfeld, D., 1978. Infant recognition of mother's voice. Percept 7 (5), 491–497.
Mitchell, R.L., Elliott, R., Barry, M., Cruttenden, A., Woodruff, P.W., 2003. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. Neuropsychologia 41, 1410–1421.
Monk, C.S., Nelson, E.E., McClure, E.B., Mogg, K., Bradley, B.P., Leibenluft, E., Blair, R.J., Chen, G., Charney, D.S., Ernst, M., Pine, D.S., 2006. Ventrolateral prefrontal cortex activation and attentional bias in response to angry faces in adolescents with generalized anxiety disorder. Am. J. Psychiatry 163 (6), 1091–1097.
Morris, J.S., Scott, S.K., Dolan, R.J., 1999. Saying it with feeling: neural responses to emotional vocalizations. Neuropsychologia 37, 1155–1163.
Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., Kojima, S., 2001. Neural substrates for recognition of familiar voices: a PET study. Neuropsychologia 39, 1047–1054.
Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J., Kojima, S., 2012. Cerebral responses to infant-directed speech and the effect of talker familiarity. Neuroimage 59 (2), 1735–1744.
Natu, V., O'Toole, A.J., 2011. The neural processing of familiar and unfamiliar faces: a review and synopsis. Br. J. Psychol. 102 (4), 726–747. https://doi.org/10.1111/j.2044-8295.2011.02053.x.
Ockleford, E.M., Vince, M.A., Layton, C., Reader, M.R., 1988. Responses of neonates to parents' and others' voices. Early Hum. Dev. 18, 27–36.
Patel, S., Scherer, K.R., Björkner, E., Sundberg, J., 2011. Mapping emotions into acoustic space: the role of voice production. Biol. Psychol. 87 (1), 93–98.
Paus, T., 2005. Mapping brain maturation and cognitive development during adolescence. Trends Cogn. Sci. 9 (2), 60–68.
Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spat. Vision. 10, 437–442.
Poldrack, R.A., Mumford, J.A., 2009. Independence in ROI analysis: where is the voodoo? Soc. Cogn. Affect. Neurosci. 4 (2), 208–213. https://doi.org/10.1093/scan/nsp011.
Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J., Vul, E., Yarkoni, T., 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat. Rev. Neurosci. 18 (2), 115.
Raschle, N.M., Smith, S.A., Zuk, J., Dauvermann, M.R., Figuccio, M.J., Gaab, N., 2014. Investigating the neural correlates of voice versus speech-sound directed information in pre-school children. PLoS One 9 (12), e115549.
Rosnow, R.L., Rosenthal, R., 2003. Effect sizes for experimenting psychologists. Can. J. Exp. Psychol. 57 (3), 221–237. https://doi.org/10.1037/h0087427.
Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P., 2005. Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. Neuroimage 28 (4), 848–858.
Saxe, R., Brett, M., Kanwisher, N., 2006. Divide and conquer: a defense of functional localizers. Neuroimage 30 (4), 1088–1096.
Scherf, K.S., Thomas, C., Doyle, J., Behrmann, M., 2013. Emerging structure–function relations in the developing face processing system. Cereb. Cortex 24 (11), 2964–2980.
Scherf, K.S., Luna, B., Avidan, G., Behrmann, M., 2011. "What" precedes "which": developmental neural tuning in face-and place-related cortex. Cereb. Cortex 21 (9), 1963–1980.
Schirmer, A., Kotz, S.A., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cogn. Sci. 10, 24–30.
Seltzer, L.J., Prososki, A.R., Ziegler, T.E., Pollak, S.D., 2012. Instant messages vs. speech: hormones and why we still need to hear each other. Evol. Hum. Behav. 33 (1), 42–45.
Semel, E., Wiig, E., Secord, W., 2003. Clinical Evaluation of Language Fundamentals −4 (CELF-4). PyschCorp, San Antonio, TX.
Shackman, J.E., Shackman, A.J., Pollak, S.D., 2007. Physical abuse amplifies attention to threat and increases anxiety in children. Emotion 7 (4), 838–852. https://doi.org/10.1037/1528-3542.7.4.838.
Shah, N.J., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., Fink, G.R., 2001. The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. Brain 124, 804–815.
Silk, J.S., Lee, K.H., Elliott, R.D., Hooley, J.M., Dahl, R.E., Barber, A., Siegle, G.J., 2017. "Mom—I don't want to hear it': brain response to maternal praise and criticism in adolescents with major depressive disorder. Soc. Cogn. Affect. Neurosci. 12 (5), 729–738.
Spinelli, M., Fasolo, M., Mesman, J., 2017. Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. Dev. Rev. 44, 1–18.
Uchida, M.O., Arimitsu, T., Yatabe, K., Ikeda, K., Takahashi, T., Minagawa, Y., 2018. Effect of mother's voice on neonatal respiratory activity and EEG delta amplitude. Dev. Psychobiol. 60 (2), 140–149.
Vannest, J.J., Karunanayaka, P.R., Altaye, M., Schmithorst, V.J., Plante, E.M., Eaton, K.J., Holland, S.K., 2009. Comparison of fMRI data from passive listening and active-response story processing tasks in children. J. Magn. Reson. Imaging 29 (4), 971–976. https://doi.org/10.1002/jmri.21694.
Voegtline, K.M., Costigan, K.A., Pater, H.A., DiPietro, J.A., 2013. Near-term fetal response to maternal spoken voice. Infant Behav. Dev. 36 (4). https://doi.org/10.1016/j.infbeh.2013.05.002.
Wang, X., Merzenich, M.M., Sameshima, K., Jenkins, W.M., 1995. Remodelling of hand representation in adult cortex determined by timing of tactile stimulation. Nature 378 (6552), 71–75.
Whalen, P.J., Johnstone, T., Somerville, L.H., Nitschke, J.B., Polis, S., Alexander, A.L., Kalin, N.H., 2008. A functional magnetic resonance imaging predictor of treatment response to venlafaxine in generalized anxiety disorder. Biol. Psychiatry 63 (9), 858–863. https://doi.org/10.1016/j.biopsych.2007.08.019.
Wildgruber, D., Ackermann, H., Kreifelts, B., Ethofer, T., 2006. Cerebral processing of linguistic and emotional prosody: fmri studies. Prog. Brain Res. 156, 249–268.
Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11 (10), 946–953.