



Age-related differences in neural activation and functional connectivity during the processing of vocal prosody in adolescence

Michele Morningstar^{1,2} · Whitney I. Mattson¹ · Joseph Venticinque¹ · Stanley Singer Jr · Bhavani Selvaraj³ · Houchun H. Hu³ · Eric E. Nelson^{1,2}

Published online: 12 September 2019
© The Psychonomic Society, Inc. 2019

Abstract

The ability to recognize others' emotions based on vocal emotional prosody follows a protracted developmental trajectory during adolescence. However, little is known about the neural mechanisms supporting this maturation. The current study investigated age-related differences in neural activation during a vocal emotion recognition (ER) task. Listeners aged 8 to 19 years old completed the vocal ER task while undergoing functional magnetic resonance imaging. The task of categorizing vocal emotional prosody elicited activation primarily in temporal and frontal areas. Age was associated with a) greater activation in regions in the superior, middle, and inferior frontal gyri, b) greater functional connectivity between the left precentral and inferior frontal gyri and regions in the bilateral insula and temporo-parietal junction, and c) greater fractional anisotropy in the superior longitudinal fasciculus, which connects frontal areas to posterior temporo-parietal regions. Many of these age-related differences in brain activation and connectivity were associated with better performance on the ER task. Increased activation in, and connectivity between, areas typically involved in language processing and social cognition may facilitate the development of vocal ER skills in adolescence.

Keywords Adolescence · Emotional prosody · Vocal · Emotion recognition · Functional connectivity · Neural activation

Adolescence is a time of intensive changes to youth's brain structure and function, cognitive capacities, and socioemotional processing (Crone & Dahl, 2012; Gogtay et al., 2004; Nelson et al., 2005). Neural maturation in adolescence is evidenced by regional reductions in gray matter and global increases in white matter (Blakemore & Choudhury, 2006; Paus et al., 2011), enhanced integrity of white matter tracts (Mohammad & Nashaat, 2017; Schmithorst et al., 2002), and heightened functional network formation (Cohen Kadosh et al., 2010; Fair et al., 2009). Alongside this pattern of neural development, marked changes in social behaviour also emerge: adolescents are increasingly oriented towards

their peers (Larson et al., 1996; Nelson et al., 2005) and begin to develop complex and nuanced social relationships (Furman & Buhrmester, 1992; Güroğlu, van den Bos, & Crone, 2014). Importantly, both the neural and behavioural maturation that occur during adolescence are thought to be shaped by individuals' unique experiences with their social environment (Nelson, 2017; Tottenham, 2014). Therefore, adolescence is often considered a sensitive period for the development of social cognitive functions (Blakemore & Mills, 2014; Crone & Dahl, 2012).

Maturation changes within the "social brain network" have been linked to the simultaneous increase in adolescents' social behaviour (Nelson, Jarcho, & Guyer, 2016) and social cognition skills (Blakemore, 2008, 2012; Burnett et al., 2011; Kilford, Garrett, & Blakemore, 2016), including emotion recognition (ER). Emotion recognition, or the ability to recognize others' emotions based on nonverbal cues (e.g., facial expressions, gestures and postures, tone of voice), is essential to social competence (Halberstadt, Denham, & Dunsmore, 2001). This skill matures throughout adolescence, presumably bolstered by increasingly sophisticated cognitive abilities and the experience-driven growth of neural networks (Nelson, 2017; Nelson et al., 2016). For instance, age-related increases in white matter and neural activity in face processing areas of

✉ Michele Morningstar
michele.morningstar@nationwidechildrens.org

¹ Center for Biobehavioral Health, Nationwide Children's Hospital, Near East Office Building, 431 S. 18th St., Columbus, OH 43205, USA

² Department of Pediatrics, The Ohio State University, Columbus, OH, USA

³ Department of Radiology, Nationwide Children's Hospital, 700 Children's Drive, Columbus, OH 43205, USA

the brain have been associated with greater accuracy in recognizing facial expressions of emotion (facial ER) in 7- to 37-year-olds (Cohen Kadosh et al., 2012).

To date, behavioural and brain-based research has primarily assessed ER development using facial expressions of emotion (Blakemore, 2008; Kilford et al., 2016), which youth can reliably interpret by early adolescence (Herba & Phillips, 2004; Kolb, Wilson, & Taylor, 1992). However, less is known about the mechanisms supporting the development of emotion recognition in other modalities, such as a speaker's tone of voice (vocal ER). Beyond the content of speech, the acoustic characteristics of vocal prosody, including the pitch, intensity levels, and temporal aspects of the voice, combine to convey important information about a speaker's emotional state or social attitudes (Banse & Scherer, 1996; Mitchell & Ross, 2013). The emotional content of prosody also necessarily unfolds over time (Liebenthal et al., 2016), which requires extensive executive functioning and processing skills to track and decode adequately (Schirmer, 2017). Perhaps relatedly, vocal ER skills have been found to follow a protracted developmental trajectory (Morningstar et al., 2019; Morningstar, Ly, Feldman, & Dirks, 2018a). Indeed, although children start to correctly label basic emotions in a speaker's voice between 6 and 9 years old (Matsumoto & Kishimoto, 1983), there is evidence for continued maturation of vocal ER through childhood (Allgood & Heaton, 2015; Doherty et al., 1999; Sauter, Panattoni, & Happé, 2013; Tonks et al., 2007). Furthermore, a handful of studies have found that adult listeners outperform 11- and 12-year-olds (Brosigole & Weisman, 1995; Chronaki et al., 2015), or even 13- to 15-year-olds (Chronaki et al., 2018; Morningstar et al., 2018a), in vocal ER tasks, indicating that this skill continues to develop at least through mid-adolescence (see Morningstar et al., 2018b for a review). However, the neural mechanisms supporting this ongoing maturation remain unknown.

Prior work with adult listeners has established a network of temporal and frontal areas involved in the perception and interpretation of vocal affective prosody (Schirmer & Kotz, 2006; Wildgruber et al., 2006). The brain model for the extraction of emotional content from prosodic cues involves the integration within the superior temporal sulcus and gyrus (STS, STG) of information from the primary auditory cortex in Heschl's gyrus (A1) and the "temporal voice area" (TVA; Belin, Zatorre, & Ahad, 2002; Belin et al., 2000; Ethofer et al., 2006b; Ethofer et al., 2012; Wiethoff et al., 2008) with subcortical structures such as the amygdala and striatum (Bach et al., 2008; Ethofer et al., 2009b). The temporal areas then project to the dorso-lateral prefrontal cortex (dlPFC; Ethofer et al., 2006a), which is implicated in the explicit interpretation of emotional intent in vocal prosody (Adolphs, Damasio, & Tranel, 2002; Alba-Ferrara et al., 2011; Wildgruber et al., 2006). However, despite extensive characterization of the neural networks involved in the processing of vocal prosody in

adults, little is known about how these brain systems develop in youth.

The current study examined developmental influences on neural activation during the processing of vocal emotional prosody. We recruited youth aged 8 to 19 years old to complete a vocal ER task while undergoing functional magnetic resonance imaging (fMRI). Our goals were to a) describe the neural correlates of processing affective prosody in youth, and b) investigate age-related changes in brain activation and neural networks during the ER task. Based on existing work with adults (Ethofer et al., 2012; Wildgruber et al., 2006), we expected that youth would recruit primary auditory processing areas in the temporal lobe (A1 and the TVA), as well as frontal regions, such as the dlPFC and inferior frontal gyrus (IFG), when tasked with attributing emotional intent to heard affective prosody. Although the TVA shows a more specialized and focalized response to nonemotional voices in adults compared with children and adolescents (Bonte et al., 2013; Bonte et al., 2016), we hypothesized that age-related changes in activation during the ER task would be primarily evident in the frontal regions responsible for the cognitive labelling of emotion, rather than in primary or secondary sensory areas (Casey et al., 2005). Furthermore, we posited that frontal regions implicated in the vocal ER task would be increasingly connected with temporal regions with age, both in terms of their functional connectivity and of the increased efficiency of the superior longitudinal fasciculus (SLF), a white matter tract linking these regions.

Method

Participants

Forty-one youth (26 females) aged 8 to 19 years old ($M = 14.00$, $SD = 3.38$) participated in the study. Participants were recruited via responses to an email advertisement distributed to employees of a large children's hospital. Exclusion criteria included the presence of devices or conditions contraindicated for MRI (e.g., braces or a pacemaker; assessed using a metal screening form), gross cognitive impairments, or developmental disorders (e.g., Turner's syndrome, autism). One participant did not complete the scanner part of the study. Participants' scaled scores for the matrix reasoning and vocabulary subtests of the Wechsler Intelligence Scale for Children (WISC)/Wechsler Adult Intelligence Scale (WAIS) were average (ranging from low average to very superior). Self-report of ethnicity indicated that 68% of the sample was Caucasian, 17% Black or African American, and 15% multiracial or other ethnicities. All participants spoke English fluently as their dominant language. The Edinburgh Handedness Inventory (Oldfield, 1971) revealed that 88% of participants were right-handed, 2% left-handed, and 10% reported no

preference. The distribution of age across our sample was symmetric (skewness = -0.17), with an approximately equal number of participants at each “age” by year (kurtosis = -1.37). Written, parental consent and written participant assent or consent was obtained before the study.

Stimuli and task

Audio stimuli were selected from a set of recordings produced by three 13-year-old community-based actors (2 females), generated with the aid of emotional vignettes (Morningstar, Dirks, & Huang, 2017). Actors spoke the same five sentences (e.g., “Why did you do that?”; “I didn’t know about it”) in five emotional tones of voice: anger, fear, happiness, sadness, and neutral. Recordings retained for the current study were chosen based on adult listeners’ ratings of their recognisability and authenticity (Morningstar et al., 2018a). Each actor contributed 25 recordings to the pool of stimuli (5 sentences \times 5 emotional categories), resulting in a total of 75 recordings. The recordings varied in duration from 0.89s to 2.03s seconds, with a mean of 1.34s.

Before MRI acquisition, youth were trained on a practice task containing audio clips of exaggerated vocalizations in a mock scanner. Participants then completed a forced-choice emotion recognition task in the MRI scanner. Each trial of the task consisted of stimulus presentation, followed by a 5-second response period during which subjects selected the speaker’s intended emotion from five labels (anger, fear, happiness, sadness, neutral). Although combining response contingency with stimulus delivery creates some difficulty in interpreting patterns of activation, we opted for this design (rather than passive listening), because we felt engaging participants in emotion categorization during stimulus delivery best captured the processes we were hoping to model. The stimuli were delivered to participants via pneumatic earbuds and responses were recorded with Lumina handheld response devices inside the scanner. Stimuli were presented in an event-related design with a jittered inter-trial interval of 1 to 8 seconds (mean 4.5s). A monitor at the head of the magnet bore was visible to subjects via a mirror mounted on the head coil: a fixation cross was shown on the screen during the inter-trial interval and the auditory stimulus, and a pictogram of response labels was shown during the rating period. The ER task was completed in three runs of approximately 6 minutes in length (25 recordings per run), presented in random order. Each run contained a pseudorandomized order of recordings from all three speakers, retaining a balanced number of recordings for each emotion and sentence.

Image acquisition

Due to equipment upgrade, MRI data were acquired on two Siemens 3 Tesla scanners running identical software, using

standard 32-channel and 64-channel head coil arrays. The imaging protocol included three-plane localizer scout images and an isotropic 3D T1-weighted anatomical scan covering the whole brain (MPRAGE). Typical imaging parameters for the MPRAGE were: 1-mm pixel dimensions, 176 sagittal slices, repetition time (TR) = 2200–2300 ms, echo time (TE) = 2.45–2.98 ms, field of view (FOV) = 248–256 mm. Subsequently, functional MRI and 64-direction diffusion tensor imaging (DTI) data were acquired with echo planar imaging (EPI) acquisitions, with a voxel size of 2.5 \times 2.5 \times 3.5–4 mm. For fMRI scans, dummy data were collected for 9.2s while a blank screen was presented to participants. Imaging parameters were: TR = 1,500 ms, TE = 30–43 ms, FOV = 240 mm. For DTI scans, $b = 0$ data were acquired with the phase-encoding axis oriented in both anterior-posterior and posterior-anterior directions to allow for subsequent post-processing steps to correct for eddy currents and geometric distortion effects. DTI parameters were: TR = 1,900–2,280 ms, TE = 62–84.2 ms, FOV = 240 mm.

Image processing

EPI images were preprocessed and analyzed in AFNI, version 18.0.11 (Cox, 1996). Functional images were corrected to the first volume, realigned to the AC/PC line, and coregistered to the T1 anatomical image. The resulting image was then normalized nonlinearly to the Talairach template. After normalization, the data were spatially smoothed with a Gaussian filter (FWHM, 6-mm kernel). Within each functional run, voxel-wise signal was scaled to a mean value of 100, and any signal values above 200 were censored. Volumes for which 10% of the voxels or more were deemed to be signal outliers or contained movement greater than 1 mm between volumes were censored before analyses.

Analysis

Behavioural accuracy Based on signal detection statistics (Pollak, Cicchetti, Hornung, & Reed, 2000), participants’ hit rates (HR; correct responses) and false alarms (FA; incorrect responses) on the ER task were combined into an estimate of sensitivity (P_r , or $HR - FA$) for each emotion category. Similar to d' (i.e., $z(HR) - z(FA)$), P_r is more appropriate when subjects’ recognition accuracy is low (Snodgrass & Corwin, 1988), which often is the case in vocal ER tasks (i.e., average of 60% for standardized voice samples produced by adult actors (Johnstone & Scherer, 2000) or 50% for youth listeners interpreting youth-produced vocal affect (Morningstar et al., 2018a)). P_r ranges from -1 to 1 , where a positive value represents more correct responses than errors (i.e., $HR > FA$), and a negative value represents more errors than correct responses (i.e., $FA > HR$). Responses that were made within 150 ms of the rating period were censored from analyses due to

physiologic implausibility. One participant's behavioural data were unavailable due to equipment error. A generalized linear model was performed to examine the effect of Emotion (within-subject, 5 levels: anger, fear, happiness, sadness, neutral) and Age in years (between-subject, continuous variable of interest) on P_r , with Gender (between-subject, 2 levels) as a control variable.

fMRI analysis Event-related response amplitudes were first estimated at the subject level. We included regressors for the presentation of the auditory stimulus (amplitude-modulated in duration) convolved the hemodynamic response function. We contrasted stimuli against an implicit baseline (all nonstimuli periods). A regressor for stimulus emotion category (5 levels) and nuisance regressors for motion (6 affine directions) and scanner drift (3rd polynomial) were also included at the subject level. For group-level analyses, the contrast images produced for each participant were fit to a multivariate model (3dMVM in AFNI; Chen, Adelman, Saad, Leibenluft, & Cox, 2014) of the effect of Emotion category (within-subject, 5 levels) and mean-centered Age in years (between-subject, continuous) on whole-brain activation, with Gender (between-subject, 2 levels) as a control variable. Within this model, we computed a *t*-statistic from a general linear test of the effect of task on activation. We also computed *F*-statistics for the main effects of Emotion and Age, and for the interaction of Emotion \times Age. We employed the false discovery rate correction factor for all fMRI analyses by combining a conservative threshold ($p < 0.001$) with a cluster-size correction (minimum size of 26 voxels). This cluster-size threshold correction was generated using the spatial autocorrelation function of 3dclustsim, based on Montecarlo simulations with study-specific smoothing estimates (Cox, Reynolds, & Taylor, 2016), with two-sided thresholding and first-nearest neighbor clustering, at $\alpha = 0.05$ and $p < 0.001$. This procedure yielded a cluster threshold of 26 voxels, which was applied to all model results. This is a well-validated approach that has been widely used in neuroimaging literature (Cox et al., 2017; Kessler, Angstadt, & Sripada, 2017). As such, results presented below are clusters of activation that were larger than 26 contiguous voxels at $p < 0.001$. Regions were identified at their center of mass using the Talairach-Tournoux atlas. Mean activation in the voxels within defined clusters was extracted for follow-up analyses relating neural activation to age or performance.

DTI analysis All DTI data were analyzed offline using the Diffusion Toolbox within FSL 6.0 (<http://www.fmrib.ox.ac.uk/fsl>; Smith et al., 2004). The standard TOPUP, EDDY, and DTIFIT routines in the toolbox were used to correct image distortions and reconstruct diffusion tensor

and fractional anisotropy (FA) maps. Voxelwise statistical analysis of the FA data was performed using tract-based spatial statistics (TBSS; Smith et al., 2006). FA images were first registered to Montreal Neurological Institute (MNI) template space using linear registration and the FLIRT routine (Jenkinson, Bannister, Brady, & Smith, 2002; Jenkinson & Smith, 2001); the registered FA images were then averaged to create a mean FA image, from which a tract skeleton was generated. After the TBSS-based FA skeleton was created, masks were generated for the left and right superior longitudinal fasciculi using the Jülich histological atlas (Eickhoff, Heim, Zilles, & Amunts, 2006; Eickhoff et al., 2005). FA values for both tracts were calculated using these masks. Data for two participants were missing due to scanner acquisition error.

Results

Behavioural accuracy

The average P_r was 0.26 (Table 1), with average HR = 0.40 and average FA = 0.14. There was a significant effect of Emotion on P_r , $F(4, 144) = 2.91$, $p = 0.02$, $\eta^2 = 0.08$. Post-hoc pairwise comparisons with Šidák corrections revealed that anger was the best recognized emotion, followed by sadness and neutral (which did not differ from one another, $p > 0.05$), and happiness and fear (which did not differ from one another, $p > 0.05$; unless otherwise specified, all emotions differed significantly from one another, all $ps < 0.05$). There also was a main effect of the continuous variable of Age, $F(1, 36) = 4.34$, $p = 0.04$, $\eta^2 = 0.11$, such that older age was associated with greater accuracy (Figure 1). Lastly, there was a main effect of Gender, $F(1, 36) = 4.57$, $p = 0.04$, $\eta^2 = 0.11$, whereby females were more accurate than males. There were no interactions between Emotion and Age or Gender ($ps > 0.05$).¹

Neuroimaging data

Activation related to task We examined the general linear test for task activation compared to baseline (Table 2; Figure 2). Activation was noted along the length of the right and left STG, extending into the IFG. Other clusters were found in the frontal lobe, including at midline in the medial frontal gyrus and bilaterally in the precentral gyrus (which may reflect motor activity associated with task response), as well as in the occipital lobe (cuneus and

¹ We discovered post-data collection that scanner pulses were occasionally spuriously encoded as “neutral” responses in the behavioural data for 19 participants, due to technical errors. To verify the reliability of our findings, all analyses relating to accuracy presented here were recomputed to exclude those participants. Results were highly consistent with those presented in text (see Supplemental Materials).

Table 1 Mean sensitivity (P_r) for all participants

Emotion category	P_r (SE)
Happiness	0.12 (0.02)
Fear	0.12 (0.02)
Anger	0.48 (0.03)
Sadness	0.30 (0.03)
Neutral	0.22 (0.03)

SE = standard error of the mean. P_r is estimated from marginal means when the covariates of Age and Gender are held constant at mean value.

lingual gyrus) and subcortical structures (thalamus and caudate). Deactivation during the task also was noted bilaterally in the inferior parietal lobule, parahippocampal gyrus, middle temporal gyrus, and postcentral gyrus (Table 2).

Effect of emotion type and age on activation during task A main effect of Emotion was noted in the postcentral gyrus. There was a main effect of Age (Table 3; Figure 3, first column) in the bilateral superior frontal gyrus at midline (B-SFG), the right middle frontal gyrus (R-MFG), the left middle frontal gyrus (L-MFG), the left precentral gyrus (L-PCG), and the left inferior frontal gyrus (L-IFG). Regression analyses revealed that age linearly predicted increased mean activation in each of the five clusters (Figure 3, second column; B-SFG: $t(37) = 5.71$, $\beta = 0.68$, $p < 0.001$; R-MFG: $t(37) = 5.95$, $\beta = 0.70$, $p < 0.001$; L-MFG: $t(37) = 5.80$, $\beta = 0.69$, $p < 0.001$; L-

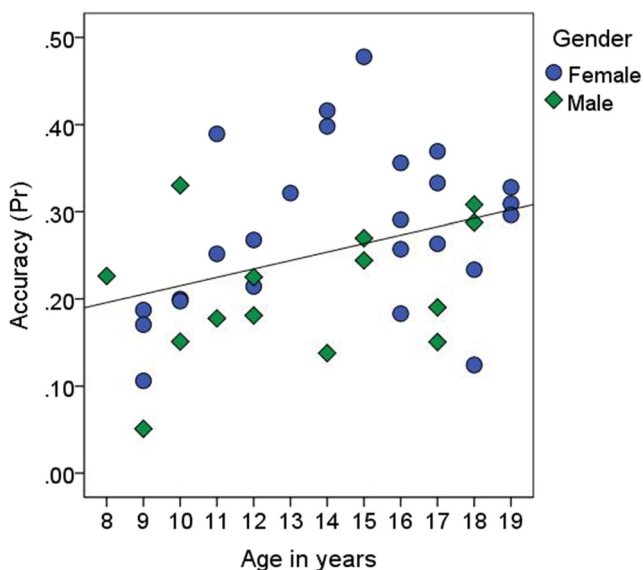


Fig. 1 Association between age (in years) and vocal ER accuracy (P_r). Female participants are identified with blue circles; male participants are identified with green triangles. The black line represents the linear relationship between the two variables ($R^2 = 0.128$).

PCG: $t(37) = 5.20$, $\beta = 0.64$, $p < 0.001$; L-IFG: $t(37) = 5.63$, $\beta = 0.67$, $p < 0.001$).²

We conducted additional regression analyses to determine whether mean activation in these age-related clusters was associated with task performance (Figure 3, third column). Greater activation in the B-SFG and L-PCG significantly predicted greater accuracy (P_r), $t(37) = 2.12$, $\beta = 0.33$, $p = 0.04$ and $t(37) = 2.12$, $\beta = 0.33$, $p = 0.04$, respectively. Activation in the other clusters was also positively related to greater P_r , although these models did not reach significance (R-MFG: $t(37) = 1.74$, $\beta = 0.27$, $p = 0.09$; L-MFG: $t(37) = 1.54$, $\beta = 0.25$, $p = 0.13$; L-IFG: $t(37) = 1.61$, $\beta = 0.26$, $p = 0.12$).

Functional connectivity related to age We conducted follow-up generalized psychophysiological interaction (gPPI) analyses (McLaren, Ries, Xu, & Johnson, 2012) to probe the functional connectivity of each age-related cluster detailed above. We first fit the same subject-level model to activation within those five regions of interest (B-SFG, R-MFG, L-MFG, L-PCG, L-IFG). We then performed a group-level model examining the effect of Age (in years, continuous variable of interest) on functional connectivity with each of those seeds. Identical cluster-size threshold corrections were applied as above (i.e., $p < 0.001$, corrected).

Age was positively associated with greater functional connection between the seed in the L-PCG and areas in the right insula (R-I), left insula (L-I), and left inferior parietal lobule/supramarginal gyrus, nearing the temporal-parietal junction (L-TPJ). Furthermore, age was also associated with increased connectivity between the seed in the L-IFG and an area in the right inferior parietal lobule/supramarginal gyrus (R-TPJ; see Table 4 and Figure 4, columns 1-3). Follow-up regression analyses (Figure 4, column 4) indicated that greater connectivity between the seed and most target regions was itself related to increased accuracy (P_r) in the ER task (R-I: $t(37) = 2.42$, $\beta = 0.37$, $p = 0.02$; L-I: $t(37) = 2.52$, $\beta = 0.38$, $p = 0.02$; R-TPJ: $t(37) = 2.53$, $\beta = 0.38$, $p = 0.02$; L-TPJ: $p = 0.11$). Thus, in addition to age-related increases in frontal activation during the vocal ER task, the strength of connections between these frontal regions and both the insula and temporal-parietal junction was also related to age and task performance.

DTI Age was associated with greater FA in both the left and right SLF (Figure 5), $t(37) = 2.55$, $\beta = 0.39$, $p = 0.02$, and $t(37) = 2.67$, $\beta = 0.40$, $p = 0.01$, respectively. FA in these tracts was positively, but not significantly, associated with task performance ($ps > 0.29$).

² Age also was positively associated with activation (t -value) at the peak voxel of each cluster, although the relationship between the two variables was attenuated: B-SFG: $t(37) = 1.59$, $\beta = 0.25$, $p = 0.12$; R-MFG: $t(37) = 2.84$, $\beta = 0.42$, $p < 0.01$; L-MFG: $t(37) = 3.58$, $\beta = 0.50$, $p < 0.001$; L-PCG: $t(37) = 1.29$, $\beta = 0.21$, $p = 0.21$; L-IFG: $t(37) = 2.16$, $\beta = 0.33$, $p < 0.05$).

Table 2 Activation during stimulus presentation compared with baseline

Structure	<i>t</i> score	<i>k</i>	<i>x</i>	<i>y</i>	<i>z</i>	Brodmann area
R superior temporal gyrus, R inferior frontal gyrus	14.77	3532	55	-10	5	22
Bilateral thalamus and caudate, L inferior frontal gyrus	5.22	2838	-14	-1	5	-
L superior temporal gyrus, L insula	13.40	2168	-53	-18	6	41
Bilateral cuneus and lingual gyrus	5.61	1952	3	-74	8	18
Bilateral medial frontal gyrus and cingulate gyrus	6.67	853	0	6	51	6
L precentral gyrus	6.75	205	-45	-6	50	6
R precentral gyrus	5.60	137	46	-8	51	6
L inferior parietal lobule and supramarginal gyrus	-4.45	854	-53	-41	36	40
R inferior parietal lobule and supramarginal gyrus	-5.98	806	54	-41	34	40
R postcentral gyrus	-5.37	424	23	-39	59	5
R parahippocampal gyrus	-8.49	338	30	-29	-9	-
L parahippocampal gyrus	-6.59	327	-28	-29	-8	-
L middle temporal gyrus, L middle occipital gyrus	-7.22	325	-48	-60	-2	19
R middle frontal gyrus, R superior frontal gyrus	-5.46	192	34	21	44	8
R insula, R lentiform nucleus	-3.80	170	35	-5	8	-
L postcentral gyrus	-4.02	154	-20	-35	60	3
R middle temporal gyrus, R middle occipital gyrus	-5.36	142	46	-61	-2	37
L superior frontal gyrus	-3.32	133	-30	56	28	9
L medial frontal gyrus	-5.10	112	-6	52	13	10

Clusters were formed using 3dclustsim at $p < 0.001$ (corrected, with cluster-size threshold of 26 voxels). Clusters of activation greater than 78 voxels (3 times the cluster threshold of 26) are presented here for brevity; all other clusters surviving thresholding of 26 voxels are available from the first author. R = right, L = left. Positive *t*-scores denote activation during the ER task compared to baseline; negative *t*-scores denote deactivation of that area. *k* = cluster size in voxels. *xyz* coordinates represent the cluster's center of mass, in Talairach-Tournoux space.

Discussion

The current study's goals were to describe the neural correlates of vocal ER in youth and to examine age-related changes in neural activation during this social cognitive task. Better performance on the ER task was associated with older age. The task of attributing emotional intent to

vocal stimuli activated temporal and frontal areas similar to those noted in work with adult listeners. We also found age-related increases in activation in several focal regions, primarily in the frontal lobe. Direct comparison of performance and brain data suggest that improvement in ER may be related to age-related increases in a) activation of frontal regions, and b) functional and tract-based

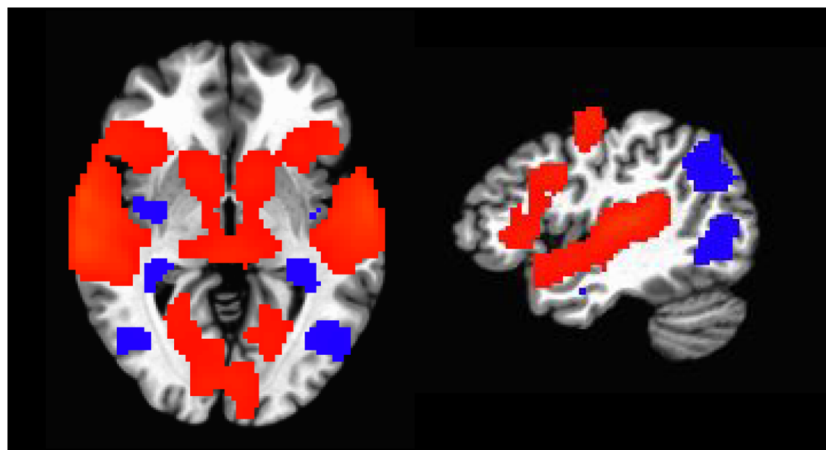


Fig. 2 Activation associated with presentation of auditory stimuli. Red areas denote increased activation during the task compared to baseline; blue denotes deactivation during the task compared to baseline. Clusters

were formed using 3dclustsim at $p < 0.001$ (corrected, with a cluster size threshold of 26 voxels). Refer to Table 2 for description of regions of activation.

Table 3 Effect of age on activation during stimulus presentation compared with baseline

Structure	<i>F</i>	<i>k</i>	<i>x</i>	<i>y</i>	<i>z</i>	Brodman area
Bilateral superior frontal gyrus (B-SFG)	24.95	85	4	10	53	6
R middle frontal gyrus (R-MFG)	28.15	64	42	20	28	9
L middle frontal gyrus (L-MFG)	21.28	41	-45	41	25	46
L precentral gyrus (L-PCG)	17.21	35	-26	-9	63	6
L inferior frontal gyrus (L-IFG)	13.85	34	-45	4	33	9

Clusters listed here represent areas in which there was a main effect of Age on activation during stimulus presentation, controlling for Emotion category and Gender in the model. Clusters were formed using 3dclustsim at $p < 0.001$ (corrected, with a cluster-size threshold of 26 voxels). R = right, L = left. *k* = cluster size in voxels. *xyz* coordinates represent cluster's center of mass, in Talairach-Tournoux space.

connectivity between frontal areas, the insula, and the temporal-parietal junction.

Emotion recognition performance

Youth's vocal ER ability was positively associated with age: specifically, older youth were more sensitive to distinctions among emotions (P_r) than younger youth across all emotions. Generally, accuracy was poorer in this sample than what is typically observed in similar developmental studies outside the scanner (i.e., equivalent to 40%, compared with 50% that has been observed with youth listeners identifying the same emotions portrayed by youth speakers; Morningstar et al., 2018a), although it remained above chance level (20%). The combination of scanner environment, logistics of button response requirements, and inclusion of a "neutral" category (Frank & Stennett, 2001) may have contributed to reduced accuracy. However, the pattern of P_r across different emotion categories is nearly identical to that noted in prior work (e.g., happiness is poorly recognized, whereas anger and sadness are well recognized; Johnstone & Scherer, 2000; Morningstar et al., 2018a), and the finding that performance is positively associated with age is consistent with previous research noting maturation of vocal ER skills across adolescence (Chronaki et al., 2015; Morningstar et al., 2019).

Neural correlates of vocal ER in youth

Across all participants, the ER task generated strong and widespread activation in brain networks previously implicated in auditory and affective processing, particularly in frontal and temporal areas. Bilateral activation was noted along the length of the STG, in the IFG near the pars opercularis, and the medial frontal gyrus. These findings are consistent with previous work on adults' neural representation of vocal prosody (Schirmer, Kotz, & Friederici, 2002; Wildgruber et al., 2006), suggesting that the current model of prosody processing is broadly applicable to children and adolescents as well. It also is noteworthy that the pattern of activation observed during the ER task occurs at least partially within areas of the "social

brain" implicated in social cognition and emotional processing, including the dlPFC and IFG (Prochnow et al., 2013; Wilson-Mendenhall, Barrett, & Barsalou, 2013), and superior temporal regions (Redcay, 2008). Although activation in some areas of the STG (the auditory cortex and TVA) may be specific to vocal processing, our findings suggest that the processes involved in vocal ER may overlap with those recruited to interpret other types of social and emotional signals, such as facial expressions of emotion (Yovel & Belin, 2013).

Age-related changes in neural activation during vocal ER

There was a positive association between age and activation in the superior frontal gyrus (SFG) at midline, the bilateral MFG, the left precentral gyrus, and the left IFG. As hypothesized, these age-related changes in activation were primarily evident in the prefrontal cortex, which is thought to mature in structure and function later in development than lower-order sensory cortices (Gogtay et al., 2004), such as the auditory cortex or temporal voice areas in the STG. Activation in the highlighted frontal areas may be related to functions that also continue to grow during adolescence, such as the top-down processing of emotion, language, and social cues. For instance, the dorsal SFG has been implicated in the regulation and reappraisal of emotion in adults (Buhle et al., 2014; Li et al., 2018). Furthermore, the MFG (BA9) has been shown to play an important role in understanding affective and linguistic prosody (meta-analysis by Belyk & Brown, 2013). The left IFG has similarly been involved in phonological, semantic, and syntactic processing (Heim, Opitz, Müller, & Friederici, 2003; Vigneau et al., 2006) and the explicit evaluation of vocal affect (Alba-Ferrara et al., 2011; Bestelmeyer et al., 2014; Frühholz, Ceravolo, & Grandjean, 2012) and is suspected to play a role in the interpretation of dynamic, temporal information (Frühholz & Grandjean, 2013; Schirmer, 2017). Outside of vocal processing, the right MFG and left precentral gyrus have been implicated in the decoding of non-verbal cues in facial expressions (Cohen Kadosh et al., 2012). Given that activation in the SFG and left precentral gyrus, and

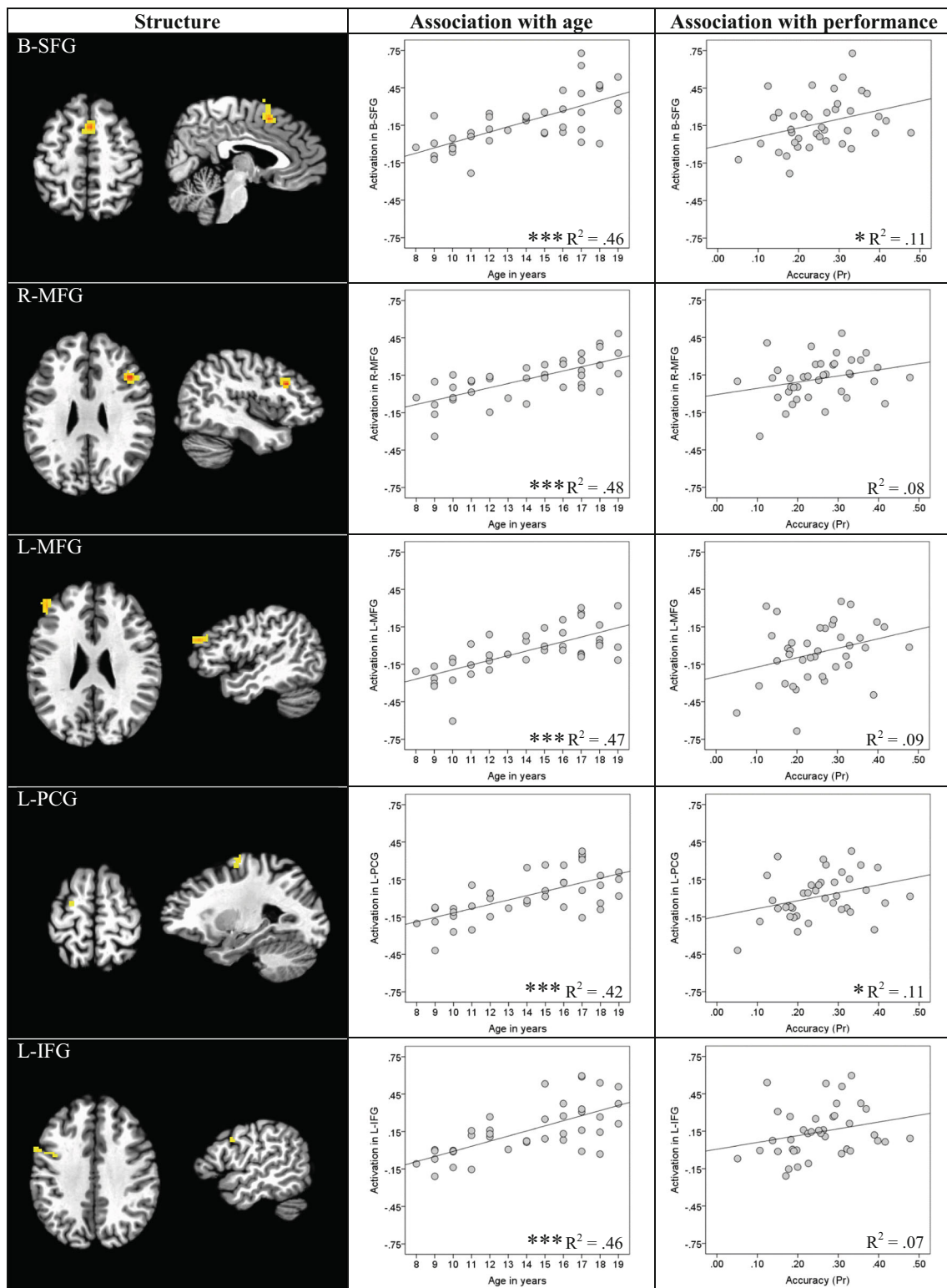


Fig. 3 Age-related changes in activation. Clusters were formed using 3dclustsim at $p < 0.001$ (corrected, with a cluster size threshold of 26 voxels). The first column illustrates five clusters of age-related activation (see Table 3 for description of each cluster). The second column contains scatterplots of the association between activation in each cluster and age. The third column contains scatterplots of the association between activation in each cluster and task performance (P_r , or sensitivity). On

each scatterplot, R^2 indicates the amount of variance in activation explained by either age or P_r ; the significance of the association between both variables is noted as * $p < 0.05$, *** $p < 0.001$. Brain images are rendered in the Talairach-Tournoux template space. L = left, R = right. SFG = superior frontal gyrus, MFG = middle frontal gyrus, PCG = precentral gyrus, IFG = inferior frontal gyrus.

Table 4 Generalized psychophysiological interaction analyses on functional connectivity with clusters of age-related activation

Structure	<i>t</i>	<i>k</i>	<i>x</i>	<i>y</i>	<i>z</i>	Brodman area
Seed in L-PCG						
R insula (R-I)	4.22	247	44	1	4	13
L insula (L-I)	5.11	43	-39	-4	1	13
L inferior parietal lobule/supramarginal gyrus (L-TPJ)	4.98	30	-56	-38	28	40
Seed in L-IFG						
R inferior parietal lobule/supramarginal gyrus (R-TPJ)	4.01	28	52	-35	30	40

Clusters listed here represent areas in which there was a main effect of Age on functional connectivity with the seed region. Clusters were formed using 3dclustsim at $p < 0.001$ (corrected, with a cluster-size threshold of 26 voxels). L-PCG = left precentral gyrus, L-IFG = left inferior frontal gyrus. R = right, L = left. *k* = cluster size in voxels. *xyz* coordinates represent the cluster's center of mass, in Talairach-Tournoux space. Some activations were noted in the cerebellum (culmen) but are not noted here (available from first author). No effects related to Age were noted for seeds in the bilateral superior frontal gyrus (B-SFG), right middle frontal gyrus (R-MFG), or left middle frontal gyrus (L-MFG).

to some extent in the right MFG, predicted increased performance on the ER task, it is possible that increased activation in these prefrontal regions supports the development of vocal emotion recognition skills with age.

The age-related changes in activation within these areas may also reflect increased efficiency and sensitivity in older adolescents compared to younger participants. Age was more robustly related to average activation across these regions than to activation at the peak of each cluster, suggesting that younger participants may have shown more widespread, but lower magnitude, response in these areas than the older participants. This interpretation would be consistent with theories of neurodevelopment and social cognition, such as the Interactive Specialization model (Johnson, Grossmann, & Cohen Kadosh, 2009), which posit specialization of function with age. Such developmental processes may be occurring in the identified prefrontal regions showing age-related differences in response to the vocal emotion recognition task, though longitudinal studies are needed to robustly assess such narrowing of function (Brown, Petersen, & Schlaggar, 2006; Durston et al., 2006).

Furthermore, the functional connectivity between the left precentral and inferior frontal gyri and the bilateral insula and temporal-parietal junction (TPJ) was positively associated with age. Diffusion tensor imaging results corroborated these findings: age was related to greater fractional anisotropy in the superior longitudinal fasciculi (SLF), white matter tracts connecting frontal and opercular regions to the parietal and temporal lobes (Taki et al., 2013). A network connecting these functionally similar areas has been implicated in the processing of vocal emotional prosody, whereby the TVA in the STG (Wildgruber et al., 2006) projects bilaterally to the IFG near the frontal operculum, and connects ipsilaterally to the inferior parietal lobule at the level of the TPJ (Ethofer et al., 2006a; Ethofer et al., 2012). The TPJ itself is considered a core area for the decoding of complex social cues (Blakemore & Mills, 2014; Redcay, 2008) and for theory of mind or “mentalizing”

skills (Mahy, Moses, & Pfeifer, 2014; Saxe & Wexler, 2005), and the insula is implicated in the evaluation of emotional salience (Phillips, Drevets, Rauch, & Lane, 2003). Thus, our findings suggest that, with age, there is an emergent network between frontal and temporal-parietal regions involved in processing linguistic and affective cues (Grecucci, Giorgetta, Bonini, & Sanfey, 2013).

Stronger functional connectivity between these areas may facilitate age-related improvements in vocal ER, a skill that requires the processing of both linguistic and emotional information. Indeed, greater connectivity amongst these regions was related to greater accuracy in the vocal ER task. The specialization of pathways and increasing connectivity in networks across development is thought to permit improvement in behavioural performance during a variety of social cognition tasks (Johnson, Grossmann, & Cohen Kadosh, 2009), such as the perception of faces, the detection of biological motion, and mentalizing (Klapwijk et al., 2013). An important contribution of our study is the addition of vocal prosody to this list of social cognitive functions that mature across adolescence and rely on fronto-posterior parietal networks.

Strengths and limitations

The current results describe a neural network that may be involved in the maturation of vocal ER skills in youth. However, since we opted to contrast activation during the ER task to a baseline containing no vocal information, we cannot conclude that our findings pertain specifically to vocal emotion rather than simply auditory processing. We chose not to contrast activation in response to emotional voices to that elicited by neutral voices, given that emotionally “neutral” stimuli also contain social information that must be decoded in the same way as other “basic” emotions and could even be perceived as negative rather than truly neutral (Lee et al.,

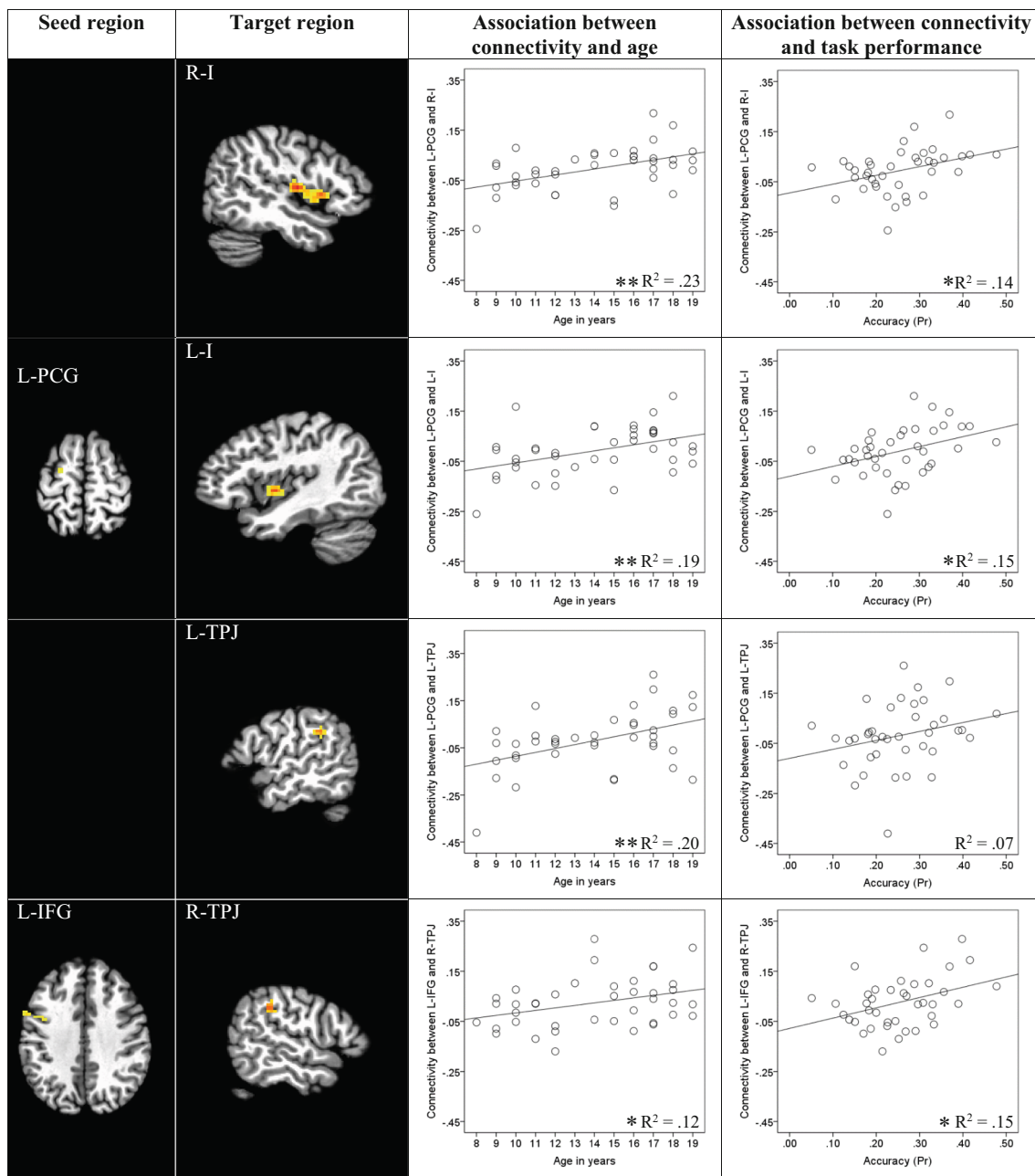


Fig. 4 Age-related changes in functional connectivity. Generalized psychophysiological interactions were computed by placing a seed in each of five clusters that showed age-related increases in activation (first column; see Figure 3 and Table 3). The second column represents target clusters for which there was an effect of Age on functional connectivity, for each seed region. Clusters were formed using 3dclustsim at $p < 0.001$ (corrected, with a cluster size threshold of 26 voxels). The third column contains scatterplots of the association between seed-target connectivity

and age. The fourth column contains scatterplots of the association between seed-target connectivity and task performance (P_r , or sensitivity). On each scatterplot, R^2 indicates the amount of variance in connectivity explained by either age or P_r ; the significance of the association between both variables is noted as * $p < 0.05$, ** $p < 0.01$. Brain images are rendered in the Talairach-Tournoux template space. L = left, R = right. I = insula, TPJ = temporal-parietal junction.

2008). Although subregions of the right STG and IFG activate more to emotional than neutral prosody during the implicit processing of vocal emotional prosody (Fruhholz et al., 2012), we were interested in understanding the processes at play in the explicit decoding of vocal affect, rather than detecting inter-emotion differences in neural activation patterns.

Despite this limitation, our findings are highly consistent with prior work and patterns of activation related specifically to listeners' task performance, suggesting that the activated regions are indeed likely to play an important functional role in vocal ER. Future work should aim to develop an adequate control for vocal prosody that retains the cognitive requirements associated with

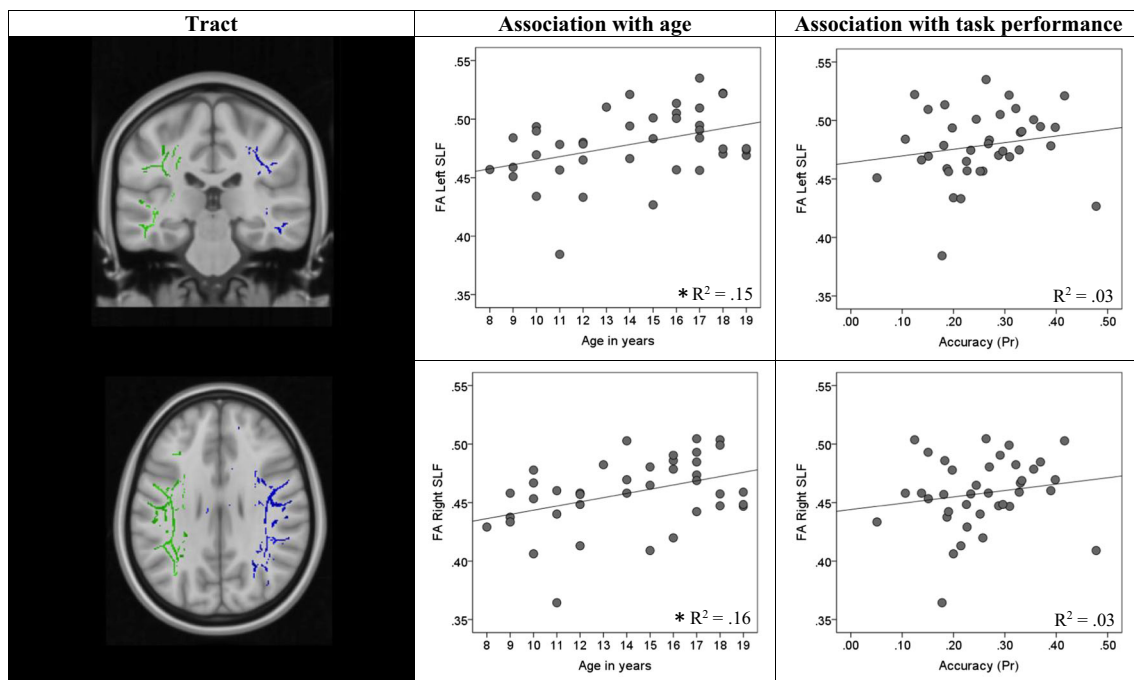


Fig. 5 Association between age and fractional anisotropy (FA) in the superior longitudinal fasciculi (SLF). Image represents the mean FA skeleton for the left (green) and right (blue) SLF across all subjects, overlaid on the 1- x 1- x 1-mm Montreal Neurological Institute template. The second column contains scatterplots representing the association

between age and FA in the left (top) and right (bottom) SLF. The third column contains scatterplots representing the association between task performance (P_r) and FA. R^2 indicates the amount of variance in FA that is explained by age; the significance of the association between both variables is noted as $* p < 0.05$.

the explicit recognition of emotional intent to confirm the specificity of the current results to the detection of vocal affect.

Of note, we did not find evidence of emotion-specific patterns of activation in this study, beyond activity in the postcentral gyrus that likely reflected motor preparation for response. Existing work with adults suggests that emotion-specific patterns of activation should be present in the temporal voice area and/or inferior frontal gyrus (Buchanan et al., 2000; Ethofer et al., 2012; Ethofer et al., 2009b; Fruhholz et al., 2012; Grandjean et al., 2005; Johnstone, van Reekum, Oakes, & Davidson, 2006; Kotz et al., 2013).³ However, whether or not discrete emotions generate unique neural signatures is a topic of considerable debate (Hamann, 2012; Lindquist & Feldman Barrett, 2012). Many studies that find such patterns of activation typically compare a limited number of emotion categories (e.g., happy vs. angry, sad vs. happy, or angry vs. neutral; Buchanan et al., 2000; Ethofer et al., 2009a; Fruhholz et al., 2012; Grandjean et al., 2005; Johnstone et al., 2006; Mitchell et al., 2003). Moreover, prior work that does include a greater number of emotion categories has suggested that emotion-specific activations may be difficult

to test empirically with functional MRI analytical approaches (Ethofer et al., 2009b; Kotz et al., 2013). Alternatively, it is possible that youth show less functional specialization in the processing of different vocal emotion patterns than would adults; though this hypothesis is strictly speculative, it would be in line with theoretical predictions of the Interactive Specialization model (Johnson et al., 2009) and previous findings regarding the development of nonemotional voice processing in the superior temporal cortex (Bonte et al., 2013; Bonte et al., 2016). Yet another possibility is that youth-produced vocal emotions, which have been found to be less distinct from one another in pitch (Morningstar et al., 2017), may not elicit as strong emotion-specific responses in the brain as would adult-produced stimuli. Future work should investigate whether emotion-specific responses to affective prosody vary depending on factors such as the age of the listener or speaker.

Moreover, the current study utilized exemplars of vocal affect embedded in speech rather than nonspeech vocalizations. Although vocalizations (such as laughs, cries, or sounds of disgust) can convey emotionality in specific circumstances, important emotional information communicated during social interactions is also contained in the prosodic variations of others' voices. Being able to parse another person's emotions or attitudes based on "the way they said something" is a crucial skill that continues to develop throughout late adolescence. Indeed, although the interpretation of vocal bursts reaches adult-like maturity

³ We do find evidence of emotion-specific activation in the left temporal voice area (superior temporal gyrus) at a less conservative threshold of $p < 0.005$, uncorrected ($k = 45$; $xyz [-44, -34, 6]$). Post-hoc analyses of the effect of emotion in this area indicate that happiness elicited more activation than all other emotions ($ps < 0.05$), and that anger elicited more activation than sadness ($p < 0.05$). Thus, expected emotion-specific patterns of activation may be present in our data, albeit weakly.

around the age of 14 to 15 years (Grosbras, Ross, & Belin, 2018), the current study and previous work have noted continued improvement beyond age 15 for the recognition of vocal affect in speech (Chronaki et al., 2018; Morningstar et al., 2018a). Thus, it is possible that age-related changes in the neural representation of these stimuli also may differ from those noted in response to speech-based vocal emotion. Additionally, the recordings used in the present study were selected based on adult ratings of their quality rather than those of adolescents. Youth may deem these exemplars as less representative of the intended emotions, which may have impacted their behavioural performance and related neural processing of these stimuli. Future work would benefit from investigating whether similar neural and behavioural responses are noted to nonspeech vocal emotion or to stimuli selected based on youth's ratings of their representativeness.

Lastly, data in the current study were cross-sectional and did not model change within individuals. Defining how brain activation and connectivity map onto behavioural development of ER by assessing change in both neural activation and task performance longitudinally within-subject will be an important future direction. The addition of an adult comparison group also would provide a more complete picture of developmental change in activation during vocal ER; future research is needed to determine the age at which youth attain adult-like maturity in the capacity to decode affective prosody. Results should be replicated in a larger sample before broader conclusions about development can be drawn.

Conclusions

The current study used a multimodal approach integrating task-based fMRI, functional connectivity, and DTI to investigate the neural mechanisms supporting the maturation of vocal ER in adolescence. This ability has been linked to social competence in youth (Nowicki & Duke, 1992, 1994) and relational satisfaction in adults (Carton, Kessler, & Pape, 1999), making it important to understand the mechanisms that facilitate its development. Our findings suggest that ongoing maturation of vocal ER in adolescence may be supported by increased connectivity between frontal and temporal-parietal areas involved in the processing of socioemotional and linguistic cues.

Funding This work was supported by internal funds at Nationwide Children's Hospital and the Fonds de recherche du Québec – Nature et technologies [grant number 207776].

Compliance with ethical standards

Competing interests The authors have no competing interests to declare.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional review board, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

- Adolphs, R., Damasio, H., & Tranel, D. (2002). Neural systems for recognition of emotional prosody: A 3-d lesion study. *Emotion, 2*(1), 23–51. doi:<https://doi.org/10.1037/1528-3542.2.1.23>
- Alba-Ferrara, L., Hausmann, M., Mitchell, R. L., & Weis, S. (2011). The neural correlates of emotional prosody comprehension: Disentangling simple from complex emotion. *PLoS One, 6*(12), e28701. doi:<https://doi.org/10.1371/journal.pone.0028701>
- Allgood, R., & Heaton, P. (2015). Developmental change and cross-domain links in vocal and musical emotion recognition performance in childhood. *British Journal of Developmental Psychology, 33*(3), 398–403. doi:<https://doi.org/10.1111/bjdp.12097>
- Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *Neuroimage, 42*(2), 919–927. <https://doi.org/10.1016/j.neuroimage.2008.05.034>
- Banase, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614.
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13*, 17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309–312.
- Belyk, M., & Brown, S. (2013). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Soc Cogn Affect Neurosci, 9*(9), 1395–1403.
- Bestelmeyer, P. E. G., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *Journal of Neuroscience, 34*(24), 8098–8105.
- Blakemore, S. J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience, 9*(4), 267–277. <https://doi.org/10.1038/nrn2353>
- Blakemore, S. J. (2012). Imaging brain development: The adolescent brain. *Neuroimage, 61*(2), 397–406. <https://doi.org/10.1016/j.neuroimage.2011.11.080>
- Blakemore, S. J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, 47*(3–4), 296–312. <https://doi.org/10.1111/j.1469-7610.2006.01611.x>
- Blakemore, S. J., & Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annu Rev Psychol, 65*, 187–207. <https://doi.org/10.1146/annurev-psych-010213-115202>
- Bonte, M., Frost, M. A., Rutten, S., Ley, A., Formisano, E., & Goebel, R. (2013). Development from childhood to adulthood increases morphological and functional inter-individual variability in the right superior temporal cortex. *Neuroimage, 83*, 739–750. <https://doi.org/10.1016/j.neuroimage.2013.07.017>
- Bonte, M., Ley, A., Scharke, W., & Formisano, E. (2016). Developmental refinement of cortical systems for speech and voice processing. *Neuroimage, 128*, 373–384. <https://doi.org/10.1016/j.neuroimage.2016.01.015>
- Brosigole, L., & Weisman, J. (1995). Mood recognition across the ages. *International Journal of Neuroscience, 82*(3–4), 169–189.
- Brown, T. T., Petersen, S. E., & Schlaggar, B. L. (2006). Does human functional brain organization shift from diffuse to focal with development? *Dev Sci, 9*(1), 9–11. <https://doi.org/10.1111/j.1467-7687.2005.00455.x>
- Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jancke, L. (2000). Recognition of emotional prosody and

- verbal components of spoken language: An fMRI study. *Cognitive Brain Research*, 9, 227–238.
- Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., . . . Ochsner, K. N. (2014). Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cereb Cortex*, 24(11), 2981–2990. <https://doi.org/10.1093/cercor/bht154>
- Burnett, S., Sebastian, C., Cohen Kadosh, K., & Blakemore, S. J. (2011). The social brain in adolescence: Evidence from functional magnetic resonance imaging and behavioural studies. *Neuroscience & Biobehavioral Reviews*, 35(8), 1654–1664. <https://doi.org/10.1016/j.neubiorev.2010.10.011>
- Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior*, 23(1), 91–100.
- Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends Cogn Sci*, 9(3), 104–110.
- Chen, G., Adelman, N. E., Saad, Z. S., Leibenluft, E., & Cox, R. W. (2014). Applications of multivariate modeling to neuroimaging group analysis: A comprehensive alternative to univariate general linear model. *Neuroimage*, 99, 571–588.
- Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. S. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology*, 33(2), 218–236.
- Chronaki, G., Wigelsworth, M., Pell, M. D., & Kotz, S. A. (2018). The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Scientific Reports*, 8(1), 8659. <https://doi.org/10.1038/s41598-018-26889-1>
- Cohen Kadosh, K., Cohen Kadosh, R., Dick, F., & Johnson, M. H. (2010). Developmental changes in effective connectivity in the emerging core face network. *Cereb Cortex*, 21(6), 1389–1394.
- Cohen Kadosh, K., Johnson, M. H., Dick, F., Cohen Kadosh, R., & Blakemore, S. J. (2012). Effects of age, task performance, and structural brain development on face processing. *Cereb Cortex*, 23(7), 1630–1642.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3), 162–173.
- Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). fMRI clustering and false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, 114(17), E3370–E3371. <https://doi.org/10.1073/pnas.1614961114>
- Cox, R. W., Reynolds, R. C., & Taylor, P. A. (2016). AFNI and clustering: False positive rates redux. *Brain Connectivity*, 7(3), 152–171.
- Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nature Reviews Neuroscience*, 13, 636–650.
- Doherty, C. P., Fitzsimons, M., Asenbauer, B., & Staunton, H. (1999). Discrimination of prosody and music by normal children. *Eur J Neurol*, 6(2), 221–226.
- Durston, S., Davidson, M. C., Tottenham, N., Galvan, A., Spicer, J., Fossella, J. A., & Casey, B. J. (2006). A shift from diffuse to focal cortical activity with development. *Dev Sci*, 9(1), 1–8. <https://doi.org/10.1111/j.1467-7687.2005.00454.x>
- Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2006). Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *Neuroimage*, 32(2), 570–582.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, 25(4), 1325–1335.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., . . . Wildgruber, D. (2006a). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *Neuroimage*, 30(2), 580–587. <https://doi.org/10.1016/j.neuroimage.2005.09.059>
- Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., . . . Wildgruber, D. (2006b). Effects of prosodic emotional intensity on activation of associative auditory cortex. *NeuroReport*, 17(3), 249–253.
- Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009a). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *J Cogn Neurosci*, 21(7), 1255–1268. <https://doi.org/10.1162/jocn.2009.21099>
- Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009b). Decoding of emotional information in voice-sensitive cortices. *Current Biology*, 19(12), 1028–1033. <https://doi.org/10.1016/j.cub.2009.04.054>
- Ethofer, T., Bretschner, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cereb Cortex*, 22(1), 191–200. <https://doi.org/10.1093/cercor/bhr113>
- Fair, D. A., Cohen, A. L., Power, J. D., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., . . . Petersen, S. E. (2009). Functional brain networks develop from a “local to distributed” organization. *PLoS Computational Biology*, 5(5), 1–14.
- Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology*, 80(1), 75.
- Fruhholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cereb Cortex*, 22(5), 1107–1117. <https://doi.org/10.1093/cercor/bhr184>
- Fruhholz, S., & Grandjean, D. (2013). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience & Biobehavioral Reviews*, 37(10), 2847–2855.
- Furman, W., & Buhrmester, D. (1992). Age and sex differences in perceptions of networks of personal relationships. *Child development*, 63(1), 103–115.
- Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., . . . Toga, A. W. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, 101(21), 8174–8179.
- Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nat Neurosci*, 8(2), 145–146. <https://doi.org/10.1038/mn1392>
- Greccucci, A., Giorgetta, C., Bonini, N., & Sanfey, A. (2013). Reappraising social emotions: The role of inferior frontal gyrus, temporo-parietal junction and insula in interpersonal emotion regulation. *Frontiers in human neuroscience*, 7, 523.
- Grosbras, M.-H., Ross, P. D., & Belin, P. (2018). Categorical emotion recognition from voice improves during childhood and adolescence. *Scientific Reports*, 8(1), 14791.
- Güroğlu, B., van den Bos, W., & Crone, E. A. (2014). Sharing and giving across adolescence: An experimental study examining the development of prosocial behavior. *Front Psychol*, 5, 1–13.
- Halberstadt, A. G., Denham, S. A., & Dunsmore, J. C. (2001). Affective social competence. *Social development*, 10(1), 79–119.
- Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. *Trends Cogn Sci*, 16(9), 458–466.
- Heim, S., Opitz, B., Müller, K., & Friederici, A. D. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Cognitive Brain Research*, 16(2), 285–296.
- Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: Behavioural and neurological perspectives. *Journal of Child*

- Psychology & Psychiatry*, 45(7), 1185–1198. <https://doi.org/10.1111/j.1469-7610.2004.00316.x>
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2), 825–841.
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2), 143–156.
- Johnson, M. H., Grossmann, T., & Cohen Kadosh, K. (2009). Mapping functional brain development: Building a social brain through interactive specialization. *Developmental Psychology*, 45(1), 151.
- Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.), *Handbook of emotions* (pp. 220–235). New York: Guilford.
- Johnstone, T., van Reekum, C. M., Oakes, T. R., & Davidson, R. J. (2006). The voice of emotion: An fMRI study of neural responses to angry and happy vocal expressions. *Soc Cogn Affect Neurosci*, 1(3), 242–249. <https://doi.org/10.1093/scan/nsl027>
- Kessler, D., Angstadt, M., & Sripada, C. S. (2017). Reevaluating “cluster failure” in fMRI using nonparametric control of the false discovery rate. *Proceedings of the National Academy of Sciences*, 114(17), E3372. <https://doi.org/10.1073/pnas.1614502114>
- Kilford, E. J., Garrett, E., & Blakemore, S. J. (2016). The development of social cognition in adolescence: An integrated perspective. *Neuroscience & Biobehavioral Reviews*, 70, 106–120. <https://doi.org/10.1016/j.neubiorev.2016.08.016>
- Klapwijk, E. T., Goddings, A. L., Burnett Heyes, S., Bird, G., Viner, R. M., & Blakemore, S. J. (2013). Increased functional connectivity with puberty in the mentalising network involved in social emotion processing. *Horm Behav*, 64(2), 314–322. <https://doi.org/10.1016/j.yhbeh.2013.03.012>
- Kolb, B., Wilson, B., & Taylor, L. (1992). Developmental changes in the recognition and comprehension of facial expression: Implications for frontal lobe function. *Brain and Cognition*, 20, 74–84.
- Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J. D. (2013). Predicting vocal emotion expressions from the human brain. *Hum Brain Mapp*, 34(8), 1971–1981. <https://doi.org/10.1002/hbm.22041>
- Larson, R. W., Richards, M. H., Moneta, G., Holmbeck, G., & Duckett, E. (1996). Changes in adolescents' daily interactions with their families from ages 10 to 18: Disengagement and transformation. *Developmental Psychology*, 32(4), 744.
- Lee, E., Kang, J. I., Park, I. H., Kim, J. J., & An, S. K. (2008). Is a neutral face really evaluated as being emotionally neutral? *Psychiatry research*, 157(1), 77–85.
- Li, F., Yin, S., Feng, P., Hu, N., Ding, C., & Chen, A. (2018). The cognitive up- and down-regulation of positive emotion: Evidence from behavior, electrophysiology, and neuroimaging. *Biol Psychol*, 136, 57–66. <https://doi.org/10.1016/j.biopsycho.2018.05.013>
- Liebethal, E., Silbersweig, D. A., & Stern, E. (2016). The language, tone and prosody of emotions: Neural substrates and dynamics of spoken-word emotion perception. *Front Neurosci*, 10, 506. <https://doi.org/10.3389/fnins.2016.00506>
- Lindquist, K. A., & Feldman Barrett, L. (2012). A functional architecture of the human brain: Emerging insights from the science of emotion. *Trends Cogn Sci*, 16(11), 533–540.
- Mahy, C. E. V., Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Dev Cogn Neurosci*, 9, 68–81.
- Matsumoto, D., & Kishimoto, H. (1983). Developmental characteristics in judgments of emotion from nonverbal vocal cues. *International Journal of Intercultural Relations*, 7(4), 415–424.
- McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *Neuroimage*, 61(4), 1277–1286.
- Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, 41(10), 1410–1421. [https://doi.org/10.1016/s0028-3932\(03\)00017-4](https://doi.org/10.1016/s0028-3932(03)00017-4)
- Mitchell, R. L. C., & Ross, E. D. (2013). Attitudinal prosody: What we know and directions for future study. *Neuroscience & Biobehavioral Reviews*, 37(3), 471–479.
- Mohammad, S. A., & Nashaat, N. H. (2017). Age-related changes of white matter association tracts in normal children throughout adulthood: A diffusion tensor tractography study. *Neuroradiology*, 59, 715–724.
- Morningstar, M., Dirks, M. A., & Huang, S. (2017). Vocal cues underlying youth and adult portrayals of socio-emotional expressions. *Journal of Nonverbal Behavior*, 41(2), 155–183. <https://doi.org/10.1007/s10919-017-0250-7>
- Morningstar, M., Dirks, M. A., Rappaport, B. I., Pine, D. S., & Nelson, E. E. (2019). Associations between anxious and depressive symptoms and the recognition of vocal socioemotional expressions in youth. *Journal of Clinical Child & Adolescent Psychology*, 48(3), 491–500. <https://doi.org/10.1080/15374416.2017.1350963>
- Morningstar, M., Ly, V. Y., Feldman, L., & Dirks, M. A. (2018a). Mid-adolescents' and adults' recognition of vocal cues of emotion and social intent: Differences by expression and speaker age. *Journal of Nonverbal Behavior*, 42(2), 237–251.
- Morningstar, M., Nelson, E. E., & Dirks, M. A. (2018b). Maturation of vocal emotion recognition: Insights from the developmental and neuroimaging literature. *Neuroscience & Biobehavioral Reviews*, 90, 221–2340. <https://doi.org/10.1016/j.neubiorev.2018.04.019>
- Nelson, E. E. (2017). Learning through the ages: How the brain adapts to the social world across development. *Cognitive Development*, 42, 84–94.
- Nelson, E. E., Jarcho, J. M., & Guyer, A. E. (2016). Social re-orientation and brain development: An expanded and updated view. *Dev Cogn Neurosci*, 17, 118–127. <https://doi.org/10.1016/j.dcn.2015.12.008>
- Nelson, E. E., Leibenluft, E., McClure, E. B., & Pine, D. S. (2005). The social re-orientation of adolescence: A neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine*, 35, 163–174.
- Nowicki, S., & Duke, M. P. (1992). The association of children's nonverbal decoding abilities with their popularity, locus of control, and academic achievement. *The Journal of genetic psychology*, 153(4), 385–393.
- Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, 18(1), 9–35.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Paus, T., Zijdenbos, A., Worsley, K., Collins, D. L., Blumenthal, J., Giedd, J. N., . . . Evans, A. C. (2011). Structural maturation of neural pathways in children and adolescents: In vivo study. *Science*, 283, 1908–1911.
- Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception i: The neural basis of normal emotion perception. *Biological Psychiatry*, 54(5), 504–514. [https://doi.org/10.1016/s0006-3223\(03\)00168-9](https://doi.org/10.1016/s0006-3223(03)00168-9)
- Pollak, S. D., Cicchetti, D., Hornung, K., & Reed, A. (2000). Recognizing emotion in faces: Developmental effects of child abuse and neglect. *Developmental Psychology*, 36(5), 679.
- Prochnow, D., Höing, B., Kleiser, R., Lindenberg, R., Wittsack, H. J., Schäfer, R., . . . Seitz, R. J. (2013). The neural correlates of affect reading: An fMRI study on faces and gestures. *Behavioural brain research*, 237, 270–277.
- Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: Implications for the

- emergence of autism. *Neuroscience & Biobehavioral Reviews*, 32(1), 123–142. <https://doi.org/10.1016/j.neubiorev.2007.06.004>
- Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31(1), 97–113. <https://doi.org/10.1111/j.2044-835X.2012.02081.x>
- Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, 43(10), 1391–1399.
- Schirmer, A. (2017). ? An ALE meta-analysis comparing vocal and facial emotion processing. *Soc Cogn Affect Neurosci*, 13(1), 1–13.
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends Cogn Sci*, 10(1), 24–30. <https://doi.org/10.1016/j.tics.2005.11.009>
- Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, 14, 228–233.
- Schmithorst, V. J., Wilke, M., Dardzinski, B. J., & Holland, S. K. (2002). Correlation of white matter diffusivity and anisotropy with age during childhood and adolescence: A cross-sectional diffusion-tensor MR imaging study. *Radiology*, 222(1), 212–218.
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., . . . Flitney, D. E. (2004). Advances in functional and structural MR image analysis and implementation as ASL. *Neuroimage*, 23, S208–S219.
- Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., . . . Matthews, P. M. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4), 1487–1505.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of experimental psychology: General*, 117(1), 34.
- Taki, Y., Thyreau, B., Hashizume, H., Sassa, Y., Takeuchi, H., Wu, K., . . . Asano, K. (2013). Linear and curvilinear correlations of brain white matter volume, fractional anisotropy, and mean diffusivity with age using voxel-based and region-of-interest analyses in 246 healthy children. *Hum Brain Mapp*, 34(8), 1842–1856.
- Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007). Assessing emotion recognition in 9–15-years olds: Preliminary analysis of abilities in reading emotion from faces, voices and eyes. *Brain injury*, 21(6), 623–629.
- Tottenham, N. (2014). The importance of early experiences for neuro-affective development. *Current Topics in Behavioral Neurosciences*, 16, 109–129. https://doi.org/10.1007/7854_2013_254
- Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., . . . Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage*, 30(4), 1414–1432.
- Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. *Neuroimage*, 39(2), 885–893. <https://doi.org/10.1016/j.neuroimage.2007.09.028>
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, 156, 249–268. [https://doi.org/10.1016/S0079-6123\(06\)56013-3](https://doi.org/10.1016/S0079-6123(06)56013-3)
- Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Situating emotional experience. *Frontiers in human neuroscience*, 7, 764.
- Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends Cogn Sci*, 17(6), 263–271. <https://doi.org/10.1016/j.tics.2013.04.004>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.