

Longitudinal change in neural response to vocal emotion in adolescence

Michele Morningstar,^{1,2,3} Whitney I. Mattson,² and Eric E. Nelson^{2,3}

¹Department of Psychology, Queen's University, Kingston, ON K7L 3L3, Canada

²Center for Biobehavioral Health, Nationwide Children's Hospital, Columbus, OH 43205, USA

³Department of Pediatrics, The Ohio State University, Columbus, OH 43205, USA

Correspondence should be addressed to Michele Morningstar, Department of Psychology, Queen's University, 62 Arch Street, Kingston, ON K7L 3L3, Canada.

E-mail: michele.morningstar@queensu.ca.

Abstract

Adolescence is associated with maturation of function within neural networks supporting the processing of social information. Previous longitudinal studies have established developmental influences on youth's neural response to facial displays of emotion. Given the increasing recognition of the importance of non-facial cues to social communication, we build on existing work by examining longitudinal change in neural response to vocal expressions of emotion in 8- to 19-year-old youth. Participants completed a vocal emotion recognition task at two timepoints (1 year apart) while undergoing functional magnetic resonance imaging. The right inferior frontal gyrus, right dorsal striatum and right precentral gyrus showed decreases in activation to emotional voices across timepoints, which may reflect focalization of response in these areas. Activation in the dorsomedial prefrontal cortex was positively associated with age but was stable across timepoints. In addition, the slope of change across visits varied as a function of participants' age in the right temporo-parietal junction (TPJ): this pattern of activation across timepoints and age may reflect ongoing specialization of function across childhood and adolescence. Decreased specialization in the striatum and TPJ across timepoints was associated with better emotion recognition accuracy. Findings suggest that specialization of function in social cognitive networks may support the growth of vocal emotion recognition skills across adolescence.

Key words: adolescence; development; fMRI; vocal emotion; emotion recognition; social cognition

Adolescence is characterized by marked changes in hormones, social behaviour, and emotional and cognitive abilities (Steinberg and Morris, 2001; Crone and Dahl, 2012; Nelson *et al.*, 2016). At a neurobiological level, the adolescent brain continues to undergo extensive maturation in structure and function throughout the teenage years (Giedd *et al.*, 1999; Paus, 2005; Blakemore, 2012). In particular, regions involved in processing social information—including areas like the medial prefrontal cortex (mPFC), superior temporal sulcus (STS), temporo-parietal junction (TPJ) and temporal pole (Carrington and Bailey, 2009; Mills *et al.*, 2014)—become increasingly specialized with age and ongoing experience with social stimuli (Johnson *et al.*, 2009). According to the Interactive Specialization model of neurodevelopment (Johnson, 2000), maturation of function in these areas of the brain may be indexed by increased efficiency of processing, more focalized activation to a narrower range of relevant stimuli and greater connectivity within the social brain network. Such patterns of change in engagement and efficiency of function are thought to support improved social cognitive abilities in adolescence, such as face processing, biological motion detection, and mentalizing or

emotion recognition (ER) (Blakemore, 2008; Johnson *et al.*, 2009; Burnett *et al.*, 2011; Kilford *et al.*, 2016).

Most work probing developmental influences on neural engagement during ER tasks has examined youth's response to emotional faces (see review by Leppänen and Nelson, 2006). However, other non-verbal cues—such as a speaker's tone of voice, beyond the content of their speech—also convey important social and affective information that must be decoded in social contexts (Banse and Scherer, 1996; Mitchell and Ross, 2013). There is substantial behavioural evidence that the capacity to recognize emotional intent in others' voices (vocal ER) is undergoing active maturation well into the teenage years (Chronaki *et al.*, 2015; Grosbras *et al.*, 2018; see review by Morningstar *et al.*, 2018b). Like with other social cognitive functions, ongoing development in vocal ER in adolescence is likely bolstered by learning-related processes. To the extent that experience shapes learning, the increased exposure to affective cues in social interactions with peers—which become highly salient and central to teenagers during adolescence (Spear, 2000)—may scaffold teenagers' vocal ER skills over time. In addition, the specialization of relevant neural networks

Received: 3 September 2021; Revised: 25 February 2022; Accepted: 21 March 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

may support increased vocal ER with age. However, little is known about the developmental changes in youth's neural processing of vocal affect.

Models of neural response to vocal emotion in the adult brain (Schirmer and Kotz, 2006) outline three stages of processing in a fronto-operculo-temporal network (Wildgruber et al., 2006; Iredale et al., 2013; Kotz et al., 2013). According to these models, acoustic information is first extracted in the primary auditory cortex within Heschl's gyrus. Representation of meaningful suprasegmental features of the auditory stream (e.g. stress and tone) is then processed within a wider band of the STS and temporal lobe more broadly (Belin et al., 2000; Wildgruber et al., 2006; Yovel and Belin, 2013). Evaluation of prosody and its emotional content then largely implicates the bilateral inferior frontal cortex (Ethofer et al., 2006, 2009; Alba-Ferrara et al., 2011; Fruhholz et al., 2012). In addition to these auditory processing areas, other regions involved in social information processing (e.g. mPFC and TPJ) likely play a role in decoding emotional intent in others' voices. Indeed, the mentalizing network has been implicated in drawing inferences about others' emotions in multi-modal stimuli (Hooker et al., 2008; Zaki et al., 2009, 2012). Previous work by our group noted age-related changes in the involvement of such social cognitive regions in a vocal ER task (Morningstar et al., 2019). In a sample of 8- to 19-year-olds, increased age was associated with (i) greater engagement of frontal areas involved in linguistic and emotional processing and (ii) greater connectivity between frontal areas and the right TPJ, when hearing vocal stimuli. These neural patterns were associated with greater vocal ER accuracy, suggesting that emerging fronto-temporo-parietal networks are relevant to growth in this social cognitive skill in adolescence. However, this work was cross-sectional, which limits inferences about neural mechanisms of behavioural change over time.

Goals and hypotheses

The current study builds on this work by examining youth's behavioural and neural responses to vocal emotional prosody in a longitudinal framework. Youth aged 8 to 19 years (sample from Morningstar et al., 2019) completed a vocal ER task while undergoing functional magnetic resonance imaging (fMRI) at two timepoints, 1 year apart. Effects of both time (visit 1 vs visit 2; within-subject variable) and age (in years; between-subject variable) on neural response to vocal emotion were investigated simultaneously in a mixed-effect model. The current study's accelerated longitudinal design affords the opportunity to examine within-subject change across time alongside between-subject change across the age range in youth's blood-oxygen-level-dependent signal during the vocal ER task. Effects of time and age may denote different developmental patterns (Figure 1). For instance, a particular region of the brain could show decreased activation to vocal emotion across timepoints (which could be attributed to practice effects and/or increased efficiency of processing with maturation across a 1 year time span) and decreased activation across age (which could indicate that the task elicits less response in that region in older youth compared to younger participants, perhaps as a function of increased experience with vocal emotion in social situations for adolescents compared to children; Figure 1, Panel I). In contrast, a region could show stability in activation across timepoints (i.e. no effect of time) but increased activation in older than younger participants (i.e. effect of age), suggesting that response in this region may not change appreciably across a 1 year interval but may increase over a longer time frame from middle childhood to late adolescence (Figure 1, Panel B). Still another

possibility is that the amount of stability in a region's activation across timepoints differs as a function of a participant's age (i.e. interaction of time × age). Although effects of time and age are likely interdependent, each of these patterns implies a different mechanism for developmental change in neural processing.

Beyond the addition of a second timepoint to investigate responses to vocal emotion longitudinally, we build upon our previous work with this sample by isolating effects related to emotional information in the voice (i.e. contrasting each emotional category to neutral, rather than to baseline—which could index activation to any auditory stimulus; see the 'Neural activation to emotional voices' section). Based on prior research on the neural mechanisms underlying social cognition and vocal ER (Kilford et al., 2016; Morningstar et al., 2019), we hypothesized that developmental change in neural response to vocal emotion (whether over time and/or over age) would be most evident in neural regions involved in later stages of auditory emotion processing involved in interpretation (e.g. inferior frontal gyrus) and mentalizing/social cognition more broadly (e.g. mPFC, TPJ and temporal pole). Given the paucity of prior research on vocal emotion processing in the developing brain, we conducted exploratory whole-brain analyses of the effect of time and age on neural response to vocal emotion. We predicted that change in neural response may be related to ER accuracy, but did not make an a priori hypothesis about the direction of such an effect.

Methods

Participants

Time 1

Forty-one youth (26 female) ages 8–19 years ($M = 14.00$, $s.d. = 3.38$) were recruited via email advertisements circulated to employees of a large children's hospital (USA). Exclusion criteria included devices or conditions contraindicated for MRI (e.g. braces) or developmental disorders (e.g. autism or Turner's syndrome). Participants' self-report of race indicated that 68% of the sample was White, 17% was Black or African American and 15% was multi-racial or of other races. The sample's age followed a sufficiently normal distribution (mean and median age = 14.00; 61% of participants aged 1 s.d. away from mean; skewness = -0.17 ; kurtosis = -1.37). One participant did not complete the task in scanner at Time 1 (see below). Thus, 40 youth were included in analyses at Time 1.

Time 2

Thirty-four youth (23 female) participated in the second visit for this study, ~ 1 year later (with 10–16 months between visits; $M = 11.90$, $s.d. = 1.86$). Attrition was due to participants moving out of state ($n = 2$), having braces or permanent metal retainers ($n = 4$) or declining to participate ($n = 1$). The sex distribution in the sample did not differ across timepoints, $\chi^2(1, N = 75) = 0.15$, $P = 0.70$. Technical errors during data collection resulted in loss of data for one participant. Thus, 33 youth were included in analyses at Time 2.

Written parental consent and written participant assent or consent was obtained before participation. All procedures were approved by the hospital Institutional Review Board.

Procedure

Youth first received instructions and a practice version of the task (using exaggerated vocalizations as stimuli) in a mock scanner.

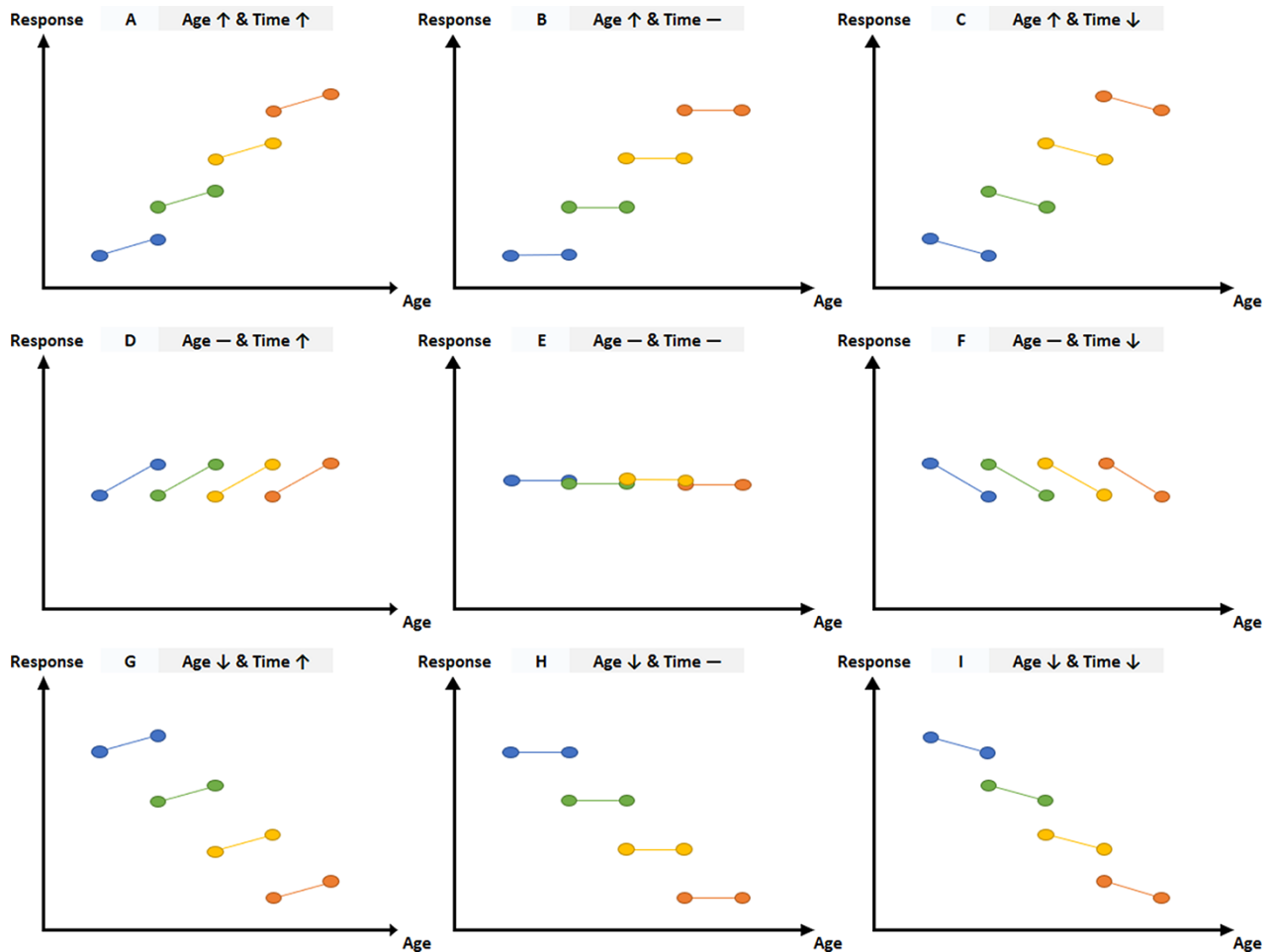


Fig. 1. Hypothetical patterns of development across age and/or time.

Note. The data depicted above are hypothetical depictions of potential patterns of development across age and/or time. The y-axis of each graph represents neural responses to stimuli and the x-axis represents participants' age. Coloured circles connected by lines represent Time 1 and Time 2 visits for a hypothetical participant.

Following this, participants completed a forced-choice vocal ER task while undergoing fMRI. Participants heard 75 recordings of vocal expressions produced by three 13-year-old actors (two females; Morningstar et al., 2017). Each actor spoke the same five sentences (e.g. 'I didn't know about it' and 'Why did you do that?') in each of five emotional tones of voice: anger, fear, happiness, sadness and a neutral expression. Recordings retained for the current study were chosen based on listeners' ratings of their recognizability and authenticity (see Morningstar et al., 2018a). Participants were asked to identify the intended emotion from five labels (anger, fear, happiness, sadness and neutral) using hand-held response boxes inside the scanner.

Each trial of the task was composed of stimulus presentation (M duration = 1.34 s, range 0.89–2.03 s), followed by a 5 s response period. Trials were presented in an event-related design with a jittered inter-trial interval of 1–8 s (mean 4.5 s). Auditory stimuli were delivered to participants via pneumatic earbuds embedded in noise-cancelling padding. A monitor at the head of the magnet bore was visible to participants via a head-coil mirror: a fixation cross was shown on the screen during the inter-trial interval and auditory stimulus delivery, and a pictogram of response labels was shown during the response period. Trials were separated into three 6 min runs (25 trials per run). Each run contained an equal number of recordings for each emotion and sentence.

Image acquisition and processing

MRI data were acquired on a Siemens 3 Tesla scanner, using a standard 64-channel head-coil array. Due to equipment upgrade during data collection, Time 1 data from five participants were acquired on a different Siemens 3T scanner with a 32-channel head coil.¹ The imaging protocol included three-plane localizer scout images, an isotropic 3D T1-weighted anatomical scan covering the whole brain (MPRAGE) and echo planar imaging (EPI) acquisitions. Imaging parameters for the MPRAGE were 1 mm voxel dimensions, 176 sagittal slices, repetition time (TR) = 2300 ms, echo time (TE) = 2.98 ms and field of view (FOV) = 248 mm. Imaging parameters for the EPI data were TR = 1500 ms, TE = 30 ms and FOV = 240 mm.

EPI images were preprocessed and analysed in Analysis of Functional NeuroImages (AFNI), version 18.1.05 (Cox, 1996). Functional images were aligned to the first volume, oriented to the anterior commissure/posterior commissure line and co-registered to the T1 anatomical image. Images then underwent non-linear warping to the Talairach template and were spatially

¹ Imaging parameters for the initial scanner were identical to those reported in the text, with the exception of the following: MPRAGE TR = 2200 ms, TE = 2.45 ms, FOV = 256 mm; EPI TE = 43 ms. Results without the five data sets acquired on a different scanner are reported in Supplemental Materials.

smoothed with a Gaussian filter (Full Width at Half Maximum, 6 mm kernel). Voxel-wise signal was scaled within-subject to a mean value of 100. Volumes in which >10% of voxels were signal outliers (above 200) or contained movement >1 mm from their subsequent volume were censored in first-level analyses. This procedure resulted in an average of 5.8% of volumes being censored across the whole sample.²

Analysis

Task performance (accuracy)

Although emotional displays can also be characterized dimensionally, we conceptualized participants' ER accuracy as the extent to which they selected the label that corresponded to the categorical emotion 'type' speakers had intended to convey. Accuracy was computed using the sensitivity index (P_r ; Corwin, 1994), a measure of accuracy based on signal detection statistics (e.g. Pollak et al., 2000). Youth's hit rates (HR; correct responses) and false alarms (FA; incorrect responses) were combined into a single estimate of sensitivity (HR – FA) for each emotion in the task.³ Similar to d' [i.e. $z(\text{HR}) - z(\text{FA})$], P_r is more appropriate for tasks in which participants' accuracy is low (Snodgrass and Corwin, 1988), as is often the case with vocal ER tasks utilizing teenage voices as stimuli (e.g. Morningstar et al., 2018a, 2019).

Linear mixed-effects models are better suited to longitudinal data than repeated-measures analysis of variance, as it does not delete participants lost to attrition list-wise and produces more accurate estimates of time-related effects for observations nested within people (Mirman, 2014). Using lmerTest (Kuznetsova et al., 2017) in R (R Core Team, 2017), we fit a linear mixed-effects model predicting P_r based on the fixed effects of Time (within-subjects variable, 2 levels: Time 1 visit vs Time 2 visit), Age (age at Time 1 visit in years; between-subjects continuous variable, mean-centred), Emotion type (within-subjects variable, 5 levels: anger, fear, happiness, sadness and neutral [sum-coded with neutral as the reference category]) and all higher-order interaction terms (Time × Age, Time × Emotion, Age × Emotion, Time × Age × Emotion). A fixed effect of Sex (between-subjects variable, 2 levels: male vs female) was included as a control variable, without interaction terms. We included a random intercept for each participant, as well as a by-participant random slope of Time (to allow change across visits to vary by participant). The equation was as follows: $P_r \sim 1 + \text{Time} \times \text{Emotion} \times \text{Age} + \text{Sex} + (1 + \text{Time} | \text{participant})$. Estimated P -values for all fixed effects were obtained from likelihood ratio tests of the full model with the effect in question against the model without it (Winter, 2013).⁴

Neural activation to emotional voices

At the subject level, regressors were included for the presentation of each type of emotional vocal stimuli (i.e. voices expressing

Table 1. P_r for each emotion type at Time 1 and Time 2

P_r	Time 1 (M, s.d.)	Time 2 (M, s.d.)
Anger	0.49 (0.20)	0.52 (0.24)
Fear	0.12 (0.12)	0.13 (0.11)
Happiness	0.12 (0.14)	0.19 (0.19)
Sadness	0.32 (0.19)	0.28 (0.19)
Neutral	0.24 (0.18)	0.35 (0.23)
Average	0.26 (0.09)	0.29 (0.12)

Note: P_r = sensitivity index (Corwin, 1994) of accuracy in the vocal ER task. M = mean, s.d. = standard deviation.

anger, fear, happiness or sadness; from stimulus onset to offset) contrasted to the neutral vocal stimuli (i.e. voices conveying a neutral expression). This model enabled us to examine both emotion-general responses (average response to all emotions vs neutral) and emotion-specific responses (e.g. happiness vs neutral, anger vs neutral, etc.). In addition to these four regressors, nuisance regressors for motion (six affine directions and their first-order derivatives) and scanner drift (within each run) were included at the subject level. At the group level, individual contrast images produced for each participant were fit to a linear mixed-effects model (3dLME in AFNI; Chen et al., 2013). The model tested effects of Time (within-subjects variable, 2 levels: Time 1 vs Time 2; with a random slope), mean-centred Age (age at Time 1 visit in years; between-subjects continuous variable, mean-centred) and Emotion type (within-subjects variable, four levels: anger vs neutral, fear vs neutral, happiness vs neutral and sadness vs neutral) on neural activation to emotional vs neutral voices, with a random intercept. All higher-order interactions were included in the model (Time × Age, Time × Emotion, Emotion × Age, Time × Age × Emotion). Sex (between-subjects variable, two levels: male vs female) was included as a control variable. Within this model, F -statistics were computed for the main effects of Time, Age and Emotion, as well as two-way interactions between these factors.

The standard false discovery rate correction factor was employed by combining a conservative cluster-forming threshold ($P < 0.001$) with a cluster-size correction. This cluster-size threshold correction was generated with AFNI's updated spatial autocorrelation function (3dclustsim with ACF (Cox et al., 2016), which addresses the correction issue raised in Eklund et al., 2016), which applies Monte Carlo simulations with study-specific smoothing estimates, two-sided thresholding and first-nearest neighbour clustering. Clusters larger than the calculated threshold of 26 contiguous voxels and comprising >25% of grey matter (based on the Talairach atlas in AFNI) are reported below. Mean activation in voxels within resulting clusters was extracted for analyses relating neural activation to ER accuracy.

Relationship between neural activation and ER accuracy

We tested the association between accuracy in the ER task (P_r) and neural response in clusters for which activation varied as a function of time and/or age. Change in neural activation in each cluster was computed as the difference between Time 1 and Time 2 for the contrast representing all emotional vs neutral voices.⁵

⁵ Difference scores can be problematic when correlations between timepoints are high (King et al., 2018). Correlations between Time 1 and Time 2 values in the current sample were significant (but small to medium in magnitude) for clusters that varied across Time (R-PCG: $r(30) = 0.44$, $P = 0.01$; R-DS: $r(30) = 0.45$, $P = 0.01$; R-IFG: $r(30) = 0.39$, $P = 0.03$) but not for other clusters (dmPFC: $P = 0.92$, R-TPJ: $P = 0.06$)—suggesting that the use of difference scores as a proxy for change across visits is acceptable.

² Age was significantly associated with proportion of censored volumes, $t(42) = -4.48$, $P < 0.001$. Censored volumes did not change across timepoints, $P = 0.25$. Although this is typical of many developmental neuroscience studies, results relating to age should be interpreted with caution.

³ P_r ranges from -1 to 1 : positive values represent more correct than incorrect responses ($\text{HR} > \text{FA}$), and negative values indicate more incorrect than correct responses ($\text{HR} < \text{FA}$). P_r is computed as follows: $[(\text{number of hits} + 0.5)/(\text{number of targets} + 1)] - [(\text{number of FA} + 0.5)/(\text{number of distractors} + 1)]$ (Corwin, 1994). Adjustments (i.e. adding 0.5 or 1) are made to prevent divisions by zero.

⁴ Due to an equipment error, behavioural data from one participant at Time 1 and one participant at Time 2 were unavailable. In addition, spurious scanner signals were found to have interfered with the software (E-Prime) used for recording participants' responses for 20 participants at Time 1 (with scanner pulses randomly coding as 'neutral' responses). To ensure that our results were not due to the erroneous encoding of certain responses, we conducted all analyses pertaining to behavioural accuracy both with and without these participants (see footnotes 4 and 5).

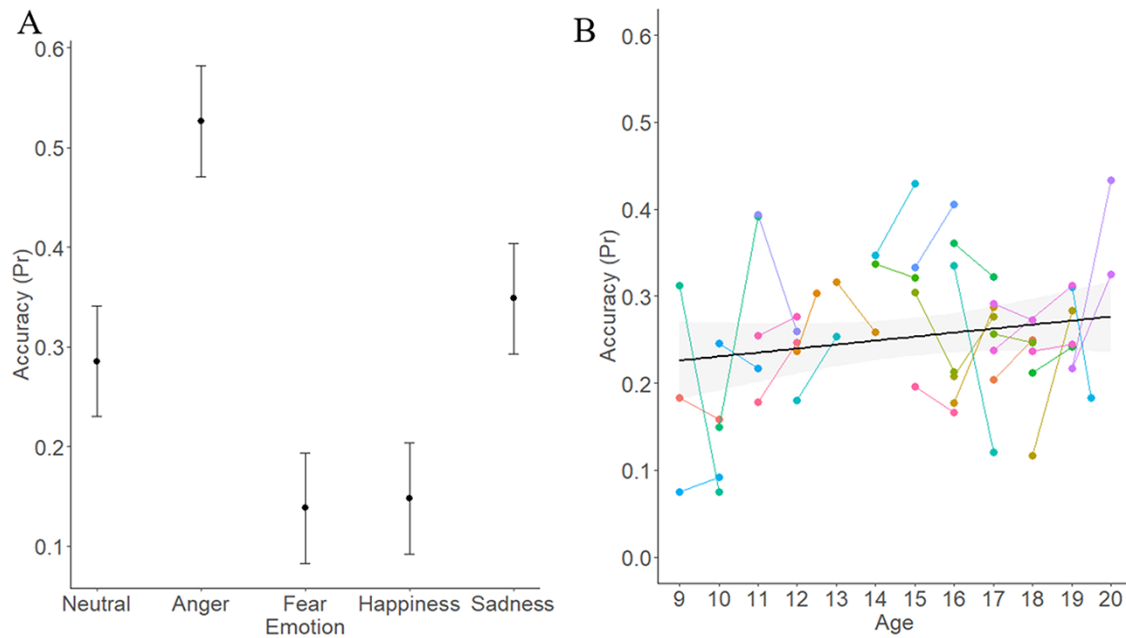


Fig. 2. Effect of Emotion and marginal effect of Age on P_r .

Note. P_r = sensitivity index (Corwin, 1994) of accuracy in the vocal ER task. Panel A represents the effect of Emotion on P_r . Panel B represents the marginal effect of Age at Time 1 on P_r . Error bars represent the standard error. Dots in Panel B represent individual participants; lines link Time 1 and Time 2 data for that participant.

An overall accuracy value for each timepoint was computed by taking the average of P_r across all emotion types at that timepoint. For each cluster, we fit a general linear model of the association between change in neural activation and P_r at Time 2, controlling for P_r at Time 1 (Rausch et al., 2003).

Results

Change in ER accuracy

Average accuracy (P_r) was 0.26 (s.d. = 0.09) at Time 1, and 0.29 (s.d. = 0.12) at Time 2 (see Table 1 for emotion-specific estimates of accuracy at both timepoints).

The linear mixed-effects model revealed a significant effect of Emotion type on accuracy ($\chi^2(16) = 209.02$, $P < 0.001$): parameter estimates suggested that anger and sadness were better recognized than other emotions on average, whereas fear and happiness were recognized less accurately than were other emotions (consistent with typical accuracy rates for emotional prosody; Johnstone and Scherer, 2000). There was also an effect of Sex ($\chi^2(1) = 6.12$, $P = 0.01$): female participants were more accurate than male participants. Lastly, there was a trend-level effect of Age on accuracy ($\chi^2(10) = 17.92$, $P = 0.056$), whereby older youth were marginally more accurate than younger youth. Other effects were not significant (all P s < 0.11). The data are illustrated in Figure 2. Parameter estimates, their standard errors and corresponding t - and P -values are included in Table 2.⁶

Change in neural activation to emotional voices

Results of the linear mixed-effects model are represented in Table 3 and Figures 3–5. There was an effect of Time in three

clusters: the right dorsal striatum (R-DS), right inferior frontal gyrus (R-IFG) and right precentral gyrus (R-PcG) showed reduced response to vocal emotions from Time 1 to Time 2 (Figure 3). There was also an effect of Age on activation in the left dorsomedial frontal gyrus at the midline (dmPFC): older youth had a greater response to emotional voices in this region than did younger participants (Figure 4). In addition, there was an interaction of Time and Age on activation in two clusters in the right TPJ (R-TPJ): younger youth showed increased response to vocal emotion in these clusters between Time 1 and Time 2, but older participants showed decreased response across timepoints (Figure 5).

Lastly, activation in several frontal and temporal regions—including, notably, the bilateral superior temporal gyri (presumed to represent the temporal voice areas, based on the location of these clusters of interest)—also varied by Emotion type (Table 3). On average, happy voices were found to elicit greater activation than emotionally neutral voices across frontal and temporal regions, whereas sad/fearful/angry voices elicited less activation than neutral voices (see the Appendix for details). Effects of Emotion type did not vary by participant age or across timepoints.

Relationship between change in neural activation to emotional voices and ER accuracy

To probe the functional consequences of changes in neural activation to emotional voices across time, we examined the association between accuracy in the ER task (P_r) and neural response in the R-DS, R-IFG, R-PcG, dmPFC and R-TPJ (i.e. clusters for which activation varied as a function of time and/or age). Results of this analysis are provided in Table 4. In all models, accuracy at Time 1 positively predicted accuracy at Time 2. In addition, changes in both the R-DS and the R-TPJ predicted accuracy at Time 2. For both clusters, a decrease in activation in these regions from Time 1 to Time 2 was associated with greater accuracy at Time 2.⁷

⁶ When the 20 participants with compromised behavioural data are removed, the effects of Sex, ($\chi^2(1) = 5.20$, $P = 0.02$) and Emotion, ($\chi^2(16) = 154.20$, $P < 0.001$), on accuracy remain significant. The effect of Age on accuracy was not significant ($P = 0.14$). Participants with and without compromised behavioural data at Time 1 did not differ in Age, $t(39) = -0.37$, $P = 0.72$.

⁷ When the 20 participants with compromised behavioural data are removed, the association between change in R-DS activation and accuracy

Table 2. Parameter estimates for linear mixed-effects model of the effect of Time, Emotion and Age on P_T

Parameter	Estimate	SE	t	P	OR [95% CI]
Intercept	0.255	0.030	8.416	<0.001 ^{***}	0.26 [0.21, 0.32]
Time	0.026	0.021	1.254	0.219	0.10 [0.03, 0.17]
Emotion Anger	0.246	0.052	4.755	<0.001 ^{***}	0.26 [0.19, 0.32]
Emotion Fear	-0.118	0.052	-2.276	0.024 [*]	-0.12 [-0.19, -0.05]
Emotion Happiness	-0.169	0.052	-3.254	0.001 ^{***}	-0.12 [-0.19, -0.05]
Emotion Sadness	0.134	0.052	2.592	0.010 ^{**}	0.08 [0.01, 0.15]
Age	0.004	0.009	0.505	0.617	0.02 [0.01, 0.04]
Sex	-0.067	0.027	-2.459	0.018 [*]	-0.07 [-0.12, -0.02]
Time × Emotion Anger	-0.010	0.034	-0.305	0.761	-0.08 [-0.19, 0.02]
Time × Emotion Fear	-0.020	0.034	-0.583	0.560	-0.09 [-0.20, 0.01]
Time × Emotion Happiness	0.030	0.034	0.882	0.379	-0.04 [-0.15, 0.06]
Time × Emotion Sadness	-0.074	0.034	-2.195	0.029 [*]	-0.15 [-0.25, -0.05]
Time × Age	0.004	0.006	0.684	0.499	-0.01 [-0.03, 0.01]
Age × Emotion Anger	0.003	0.016	0.173	0.863	-0.01 [-0.03, 0.01]
Age × Emotion Fear	-0.014	0.016	-0.903	0.367	-0.03 [-0.05, -0.01]
Age × Emotion Happiness	-0.010	0.016	-0.649	0.517	-0.02 [-0.04, <0.01]
Age × Emotion Sadness	-0.009	0.016	-0.565	0.572	-0.02 [-0.04, <0.01]
Time × Age × Emotion Anger	-0.002	0.010	-0.147	0.883	0.01 [-0.02, 0.04]
Time × Age × Emotion Fear	0.001	0.010	0.063	0.950	0.02 [-0.02, 0.05]
Time × Age × Emotion Happiness	0.007	0.010	0.705	0.482	0.02 [-0.01, 0.05]
Time × Age × Emotion Sadness	0.008	0.010	0.792	0.429	0.02 [-0.01, 0.05]

Note: P_T = sensitivity index (Corwin, 1994) of accuracy in the vocal ER task. Details of the model are provided in the text. Estimates are derived from the model fit with REML (restricted maximum likelihood); t-tests and associated P-values are derived using Satterthwaite's approximation method (as recommended by Luke, 2017). Age represents participants' chronological age (in years) at Time 1. Emotion type is sum-coded, with neutral as the reference category. Parameters referring to a specific emotion are thus representing estimates for that emotion compared to the grand average of all emotions. SE = standard error. OR = odds ratio, with 95% confidence interval (CI).

* $P < 0.05$.

** $P < 0.01$.

*** $P < 0.001$.

Table 3. Effect of Time, Emotion and Age on activation to emotional voices

Effect	Structure	F	k	x	y	z	Brodmann area
Time	R dorsal striatum (R-DS)	19.58	30	29	-14	1	n/a
	R inferior frontal gyrus (R-IFG)	15.68	27	56	19	14	44
	R precentral gyrus (R-PcG)	22.13	38	49	-11	46	4
Emotion	L postcentral gyrus	10.47	151	-39	-29	54	1
	L superior and medial frontal gyrus	11.03	128	-4	11	51	6
	L middle frontal gyrus	8.50	72	-39	16	31	8
	R superior temporal gyrus	10.80	58	59	-4	1	22
	L superior temporal gyrus	7.98	58	-51	-11	4	41
	R postcentral gyrus	7.64	46	36	-26	49	4
	L cuneus	9.69	42	-26	-74	16	19
	L lingual gyrus	8.15	38	-9	-84	-6	18
	L inferior frontal gyrus	7.10	26	-49	26	11	45
Age	L medial frontal gyrus (dmPFC)	18.84	42	-6	51	39	9
	R cingulate gyrus	22.74	33	16	16	39	8
Sex	R middle frontal gyrus	22.10	27	41	-4	44	6
	R supramarginal gyrus (R-TPJ)	29.67	41	56	-46	36	39
Time × Age	R supramarginal gyrus (R-TPJ)	29.67	41	56	-46	36	39
	R supramarginal gyrus (R-TPJ)	38.87	38	61	-51	21	39

Note: Clusters listed here represent areas in which there was a main effect of either Time, Emotion, Age at Time 1, or an interaction of Time × Age on activation to emotional voices (happiness, anger, fear and sadness) vs neutral voices. Sex is included in the model as a covariate: effects of Sex revealed that males showed less response to emotional voices than to neutral voices in the right cingulate gyrus and middle frontal gyrus—but that females did not (details available from the first author). An interaction of Time × Age was also found in the cerebellum. 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 each cluster's peak, in Talairach–Tournoux space.

Discussion

The current study is a longitudinal examination of children and adolescents' neural and behavioural responses to vocal emotional information. We investigated effects of time (within-subject change between visit 1 and visit 2) and age (between-subject

differences in activation by chronological age at scan) on 8- to 19-year-olds' neural processing of, and capacity to identify the intended emotion in, vocal expressions produced by other teenagers at two timepoints, 1 year apart. Reduced activation to vocal emotion across timepoints was noted in the R-IFG—a region involved in mentalizing tasks and the identification of emotional meaning in paralinguistic information—and the R-DS. Moreover, there was an effect of age on dmPFC activation, whereby older youth engaged this region more in response to

at Time 2 is marginally significant ($P = 0.057$). Change in R-TPJ continues to predict accuracy at Time 2 ($P = 0.043$).

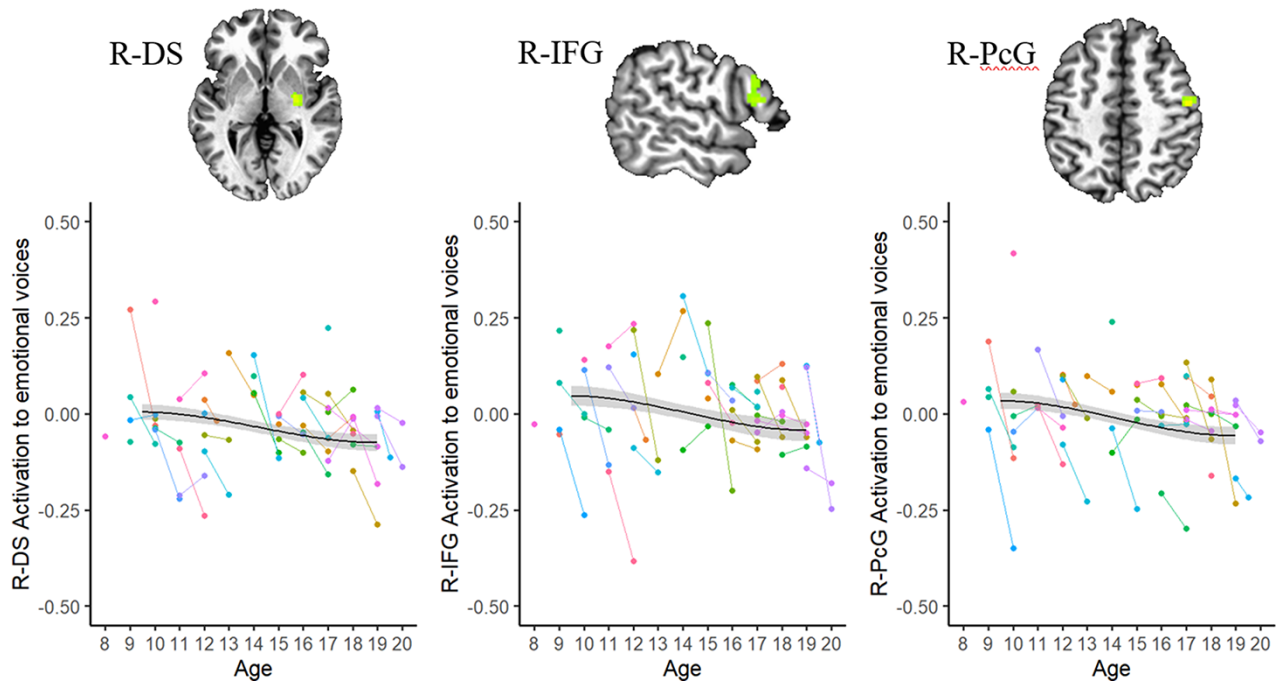


Fig. 3. Effect of Time on activation to emotional voices.

Note. R = right. DS = dorsal striatum; IFG = inferior frontal gyrus; PcG = precentral gyrus. Dots represent individual participants; lines link Time 1 and Time 2 data for that participant. Brain images are rendered in the Talairach–Tournoix template space. Clusters were formed using 3dclustsim at $P < 0.001$ (corrected, with a cluster-size threshold of 26 voxels). Refer to Table 3 for a description of regions of activation.

emotional voices than younger participants did. Lastly, younger participants showed increased right TPJ activation across time, but older participants showed the opposite pattern (interaction of time \times age). Decreased striatal and TPJ responses to vocal emotion across timepoints were also related to increased vocal ER accuracy, when controlling for baseline performance at Time 1. Although extension of this work in a larger sample and across more timepoints is needed, these findings are preliminary evidence that changes in engagement and efficiency of brain regions involved in mentalizing may support increased social cognitive capacity in interpreting vocal affect in adolescence.

As in prior investigations of change in ER skills over a 1 year period (e.g. Overgaauw et al., 2014; Taylor et al., 2015), we did not find a significant change in task accuracy across timepoints. A marginal effect of age on performance suggested that older youth were somewhat better at the task than were younger participants, which is consistent with reports of age-related change in vocal ER skills throughout adolescence (e.g. Grosbras et al., 2018; Morningstar et al., 2018a; Amorim et al., 2019). It is possible that incremental growth in vocal ER ability occurs on a longer time scale than a 1 year interval and instead matures at a more protracted rate across childhood and adolescence; however, larger samples are needed to confirm this finding.

In contrast, neural engagement with emotional voices did change over time in a 1 year period. Specifically, the right DS and IFG showed a decrease in activation across timepoints. Our findings complement prior reports of the IFG showing age-related decreases in activation during a mentalizing/facial ER task in 12- to 19-year-olds (Gunther Moor et al., 2011; Overgaauw et al., 2014). Similarly, children aged 9–14 years showed more (left) IFG activation than did adults in a mentalizing task (Wang et al., 2006). Beyond its involvement in mentalizing and theory of mind tasks (e.g. Shamay-Tsoory et al., 2009; Dricu and Frühholz, 2016), the IFG is also heavily involved in processing emotional intent in

non-verbal stimuli (Nakamura et al., 1999; Sergerie et al., 2005; Frühholz and Grandjean, 2013; Mitchell and Phillips, 2015), multi-modal social stimuli (e.g. self vs other faces and voices; Kaplan et al., 2008) and paralinguistic information in the human voice (Wildgruber et al., 2006; Kotz et al., 2013). A similar response pattern was noted in the right DS. The DS has previously been shown to be relevant to processing social reward (Guyer et al., 2012) and in coding the value of different outcomes in learning contexts (Delgado et al., 2003), which could plausibly be involved in the current task of decoding intent in social cues. However, the striatum (amongst other subcortical structures) has also been implicated in the second stage of vocal emotion processing, in which information about salience, affect and context are integrated with basic features of the auditory signal (e.g. Schirmer and Kotz, 2006; Abrams et al., 2013).

Increased efficiency of processing may account for this pattern of change in the IFG and DS: for instance, increasingly focalized response in these regions could yield lower estimates of activation across these clusters over time. This pattern of change across visits, in the context of stability across age, may be related to learning-related processes (e.g. ‘practice’ effects) that manifest in a similar way in children and adolescents. Stimuli that have been encountered before—either because they were seen in visit 1 or because youth are gaining experience with emotional expressions more broadly—may require less engagement from regions involved in later stages of auditory processing and the interpretation of social information. In conjunction with this potential specialization of function, it is also possible that lower estimates of activation in a region over time are due to spatial localization of cortical response (Johnson et al., 2009)—such that estimates of average activation in a cluster of interest may be ‘diluted’ if response occurs within a narrower set of voxels in that region. Such focalization effects likely play an important role in the developmental trajectory of vocal ER skills; indeed, decreased

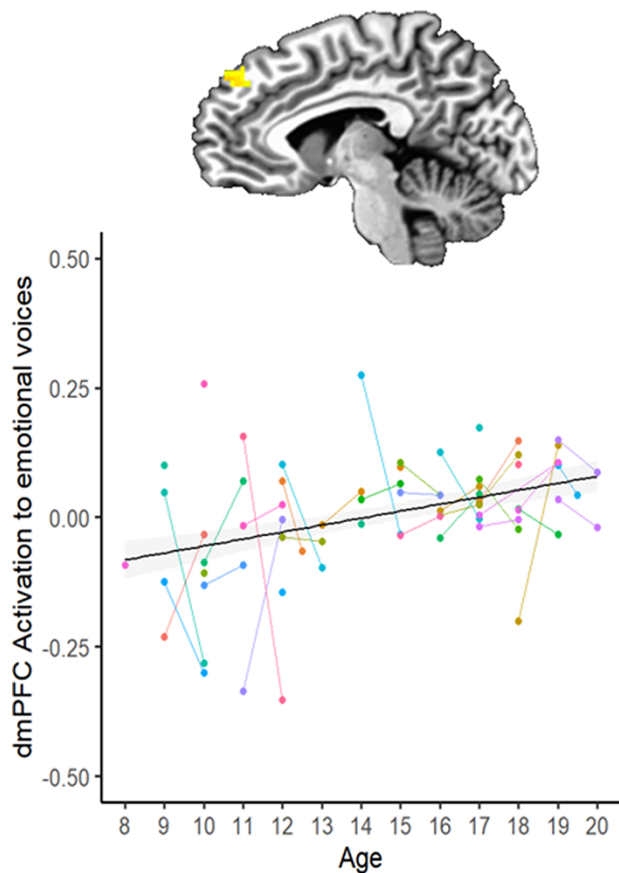


Fig. 4. Effect of Age (at Time 1) on activation to emotional voices.

Note. dmPFC = dorsomedial prefrontal cortex. Dots represent individual participants; lines link Time 1 and Time 2 data for that participant. The brain image is rendered in the Talairach–Tourmoux template space. Clusters were formed using 3dclustsim at $P < 0.001$ (corrected, with a cluster-size threshold of 26 voxels). Refer to Table 3 for a description of regions of activation.

DS activation over time predicted increased ER accuracy at Time 2 (even when controlling for baseline differences in accuracy at Time 1).

In comparison, we found that activation in the dmPFC was positively associated with age but was stable across timepoints. This pattern of change is suggestive of maturational changes that occur across a longer timescale in childhood and adolescence. As age is a proxy for a vast array of developmental influences, increased dmPFC activation with age may reflect processes that undergo a dramatic change in adolescence—including experience with peers (Nelson et al., 2016), social cognitive abilities that rely on the mPFC and other areas of the ‘social brain’ (Kilford et al., 2016) or other aspects of social information processing that are refined during the teenage years (Nelson et al., 2005). An increase in dmPFC response to vocal emotion is broadly consistent with cross-sectional age-related changes in this sample that have been previously reported (Morningstar et al., 2019). Extending our previous work that found age-related increases in response to vocal stimuli in frontal regions involved in linguistic processing and emotion categorization, the current analyses locate changes in response to emotional vocal information to a more rostral area of the dmPFC that is implicated with mentalizing (Frith and Frith, 2006). Prior work has suggested that dmPFC activation in a variety of social cognitive tasks is higher in children (Wang et al., 2006; Kobayashi et al., 2007) and adolescents (Blakemore et al., 2007; Burnett et al., 2009; Sebastian et al., 2011) than adults. However, in a longitudinal examination of

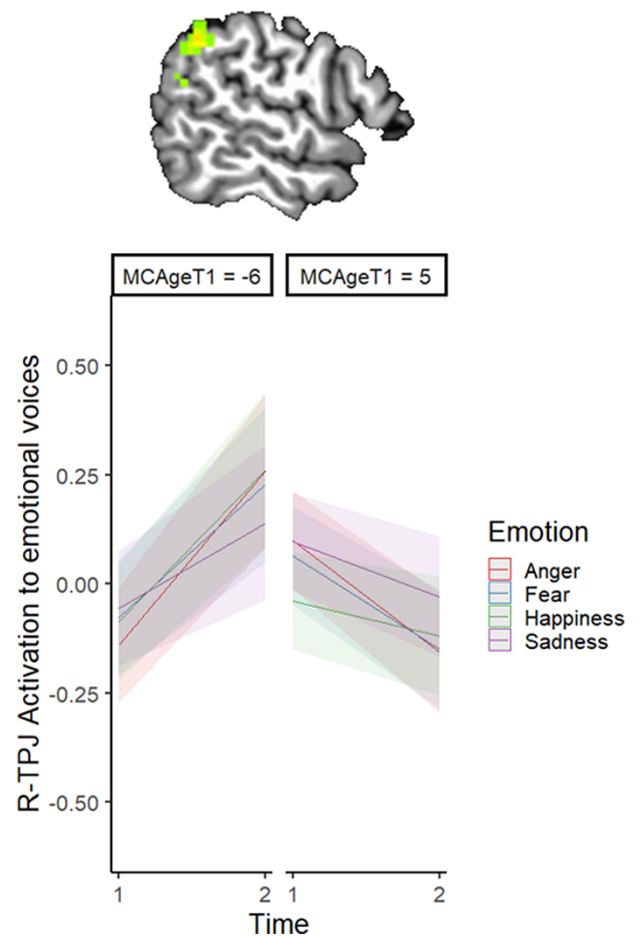


Fig. 5. Interaction of Time and Age (at Time 1) on activation to emotional voices.

Note. MCAgeT1 represents mean-centred age at Time 1. The left panel thus represents the effect of Time on R-TPJ activation to emotional voices in younger participants, whereas the right panel represents the same effect in older participants. The brain image is rendered in the Talairach–Tourmoux template space. Clusters were formed using 3dclustsim at $P < 0.001$ (corrected, with a cluster-size threshold of 26 voxels). Refer to Table 3 for a description of regions of activation.

12- to 19-year-olds’ interpretation of emotional eyes (Overgaauw et al., 2014), the same region of the dmPFC showed a curvilinear pattern of change across age, with lowest activation noted in mid-adolescents (~15–16 years old)—and stability over time (i.e. no effect of Time, as in the current investigation). Explanations for these divergent findings can only be speculative. It is possible that maturational influences on dmPFC engagement in social cognitive tasks vary as a function of stimulus. Indeed, many of the studies that report decreases in dmPFC activation across age groups (e.g. Wang et al., 2006; Kobayashi et al., 2007; Burnett et al., 2009) utilize tasks that tap into higher-order mentalizing skills (e.g. false belief tasks or detection of irony). In contrast, this developmental pattern is less consistent in studies that assess ER or the passive viewing of affective facial displays (e.g. Pfeifer et al., 2011; Overgaauw et al., 2014). Moreover, given the greater computational requirements of decoding vocal emotion—such as the need to track temporal information across a sentence (Liebenthal et al., 2016) and increased task difficulty compared to facial ER (Scherer, 2003)—the developmental trajectory of dmPFC response to emotional voices may be different from other forms of social stimuli. This is an empirical question that future research comparing neural response to various non-verbal modalities in a longitudinal framework could help answer.

Table 4. Relationship between change in neural activation and P_r at Time 2

Cluster of interest	Effect	df	F	P	η^2
R-DS	Change in neural activation	1, 27	4.618	0.041*	0.146
	Time 1 P_r	1, 27	9.829	0.004**	0.267
R-IFG	Change in neural activation	1, 27	0.591	0.449	0.021
	Time 1 P_r	1, 27	7.247	0.012*	0.212
R-PcG	Change in neural activation	1, 27	0.265	0.611	0.010
	Time 1 P_r	1, 27	8.673	0.007**	0.243
dmPFC	Change in neural activation	1, 27	0.218	0.644	0.008
	Time 1 P_r	1, 27	7.721	0.010**	0.222
R-TPJ	Change in neural activation	1, 27	5.235	0.030*	0.162
	Time 1 P_r	1, 27	5.805	0.023*	0.177

Note: P_r = sensitivity index (Corwin, 1994) of accuracy in the vocal ER task. R = right. DS = dorsal striatum, IFG = inferior frontal gyrus, PcG = precentral gyrus, dmPFC = dorsomedial prefrontal gyrus, TPJ = temporo-parietal junction. Models (described in the text) are testing the association between change in neural activation within the cluster of interest and P_r at Time 2, controlling for P_r at Time 1. df = degrees of freedom, η^2 = partial eta squared.

* $P < 0.05$.

** $P < 0.01$.

Further, change in activation within the right TPJ was dependent on both time and age. The slope of within-person change across visits varied as a function of participants' age, such that TPJ response to emotional voices increased across time for younger participants but *decreased* for older youth. The TPJ is a key region for mentalizing ability (Saxe and Kanwisher, 2003; Saxe and Wexler, 2005) and the decoding of complex social cues (Blakemore, 2008; Redcay, 2008), such as affective tones of voice. The pattern of response in this region is consistent with predictions from the Interactive Specialization model of functional brain development (Johnson, 2000), which predicts that initial stages of development are characterized by broad and diffuse patterns of activation to stimuli—but that experience and learning lead to narrowing and focalization of response over time. Such a pattern may be exemplified by the TPJ response to emotional voices. Although this region may initially show increased engagement in parsing this type of socio-emotional information in early adolescents, increased efficiency of processing (or narrowing of response localization) over time may result in lesser engagement—and potentially stronger integration to a distributed network of regions that conjointly process the information—in older youth. Indeed, decreased activation in the right TPJ across time was associated with better ER performance. These results suggest that reduced activation across time could be representative of a specialization process, which may be supporting increased social cognitive capacity in interpreting vocal affect in adolescence.

Lastly, it is noteworthy that findings of changing response across age or timepoints were primarily found in regions typically associated with social cognition, rather than auditory processing. Although the IFG is arguably involved in parsing both socio-emotional and linguistic information (Frühholz and Grandjean, 2013; Kotz et al., 2013; Mitchell and Phillips, 2015; Dricu and Frühholz, 2016; Kirby and Robinson, 2017), the TPJ and dmPFC are considered part of the default mode and mentalizing networks and contribute to the processing of social stimuli (e.g. Li et al., 2014; Meyer, 2019). As such, our findings may reflect maturational patterns in the broader default mode network. In contrast, although emotion-specific responses were noted in these regions, there were no developmental influences on activation in the auditory cortex or superior temporal gyrus—areas that are theorized to support early processing of vocal stimuli (e.g. Schirmer and Kotz, 2006). [Consistent with previous work with other non-facial non-verbal cues (e.g. Peelen et al., 2007; Ross et al., 2019), we found emotion-specific modulation of neural response in voice-sensitive regions of the brain but no evidence of an age \times emotion

interaction.] Our focus on responses to *emotional* information in the voice (e.g. contrasting emotional voices to a baseline of neutral voices) may have highlighted developmental trends in the 'social brain' areas over auditory processing regions. Nonetheless, these findings suggest the possibility that the development of the capacity to decode vocal cues of emotion in others' affective prosody may primarily rely on the maturation and fine-tuning of social cognitive, rather than early auditory processing, networks.

Strengths and limitations

The current study provides novel information about developmental influences on neural and behavioural responses to vocal emotion, an understudied aspect of social cognition. The coupling of fMRI and assessments of ER performance in a longitudinal design enables the investigation of shifts in both functional activation and associated behaviour (Schriber and Guyer, 2016). Our findings highlight potential neural mechanisms of specialization through which growth in social cognitive performance may be facilitated in adolescence. Continued efforts to delineate these developmental processes will be valuable for understanding the possible social impact of deviations to these patterns in youth who show differential responses to vocal affect (e.g. children with autism spectrum disorder; Abrams et al., 2013, 2019).

However, limitations must be noted. First, the use of an intensive methodology like fMRI by necessity limited our sample size, which is modest for an investigation of age-related change across time. Although larger longitudinal data sets are now increasingly available through consortium efforts, these typically do not include tasks that probe non-facial processing. Our findings are preliminary evidence that important developmental changes in response to non-facial non-verbal cues may be occurring during childhood and adolescence; as such, our results encourage the inclusion of tasks probing these social cognitive capacities (such as vocal ER tasks) in large-scale neuroimaging studies on socio-emotional development. Second, the inclusion of additional timepoints would permit the evaluation of non-linear patterns of change over time and allow investigations of the mechanisms behind shifts in the directionality of neural response within-person across development (e.g. in the TPJ). Third, the current study cannot disentangle effects of age from that of pubertal maturation. Variations in adrenarcheal or gonadal hormones are known to influence neural response to social stimuli (e.g. Forbes et al., 2011; Moore et al., 2012; Klapwijk et al., 2013). Although age and pubertal status are highly correlated, assessing the relationship between pubertal development and the neural processing

of vocal emotion would add to our understanding of normative development in adolescence.

It should also be noted that our analyses use neutral prosody as a comparison point to canonical emotional prosody. There is evidence that neutral emotional expressions are not necessarily equivalent to 'null' stimuli and may be perceived as negative (E. Lee et al., 2008) rather than devoid of emotion. However, using an implicit baseline (e.g. the fixation cross, or non-stimuli periods) also poses interpretative challenges; therefore, it was necessary to select one expression type as the reference level in our analyses. Given precedent for this approach (e.g. Kotz et al., 2013; meta-analysis by Schirmer, 2018), we selected neutral as a comparison point. Accuracy for neutral expressions was closest to the 'grand mean' accuracy across emotion types; moreover, response times were similar for neutral as for other expressions (see Supplemental Materials for more details). These patterns suggest that the recognition of neutral voices was likely of a similar 'difficulty level' as for other emotion types, and an adequate reference point in our analyses. However, to the extent that neutral may be more 'ambiguous' than other emotion types (and classification may be more challenging), future research would benefit from contrasting neutral to other low-intensity expressions to track developmental changes in the perceptual and neural differentiation of ambiguous non-verbal cues with age (Tahmasebi et al., 2012; Lee et al., 2020). Including a non-emotional auditory control condition in the task would also help interpret age-related changes in neural response to neutral prosodic stimuli (see Supplemental Material).

Conclusions

The current study examined changes in neural and behavioural responses to vocal emotion in adolescents. Several regions of the mentalizing network (inferior frontal gyrus, dorsomedial prefrontal cortex and TPJ) showed a change in activation over time and/or age. Notably, change in the right TPJ across timepoints varied as a function of participants' age, suggesting an 'inverted U'-like pattern of response across development. Reduced activation in the TPJ and in the DS over time was associated with better performance on the vocal ER task. Our results suggest that specialization in engagement of social cognitive networks supports the growth of vocal ER skills across adolescence. This social cognitive capacity is crucial to our understanding of the world around us, especially in modern contexts in which face-to-face interaction is more limited or in which facial communication is impaired (e.g. due to face coverings). Deepening knowledge about typical developmental trajectories of neural and behavioural responses to non-verbal social information will be crucial to understanding deviations from these norms in neurodevelopmental disorders.

Acknowledgements

We are grateful for the help of Joseph Venticinque, Stanley Singer, Jr., Roberto French, Connor Grannis, Andy Hung, Brooke Fuller and Meika Travis in collecting and processing the data, as well as to the participants and their families for their time.

Funding

This work was supported by internal funds in the Research Institute at Nationwide Children's Hospital and the Fonds de recherche du Québec—Nature et technologies (grant number 207776).

Conflict of interest

None declared.

Supplementary data

Supplementary data is available at SCAN online.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the 1964 Declaration of Helsinki and its later amendments.

References

- Abrams, D.A., Lynch, C.J., Cheng, K.M., et al. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences*, **110**(29), 12060–5.
- Abrams, D.A., Padmanabhan, A., Chen, T., et al. (2019). Impaired voice processing in reward and salience circuits predicts social communication in children with autism. *Elife*, **8**, 1–33.
- 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.
- Amorim, M., Anikin, A., Mendes, A.J., Lima, C.F., Kotz, S.A., Pinheiro, A.P. (2019). Changes in vocal emotion recognition across the life span. *Emotion*, **21**(2), 315–25.
- Banse, R., Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, **70**(3), 614–36.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, **403**, 309–12.
- Blakemore, S.J., den Ouden, H., Choudhury, S., Frith, C. (2007). Adolescent development of the neural circuitry for thinking about intentions.
- Blakemore, S.J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, **9**(4), 267–77.
- Blakemore, S.J. (2012). Imaging brain development: the adolescent brain. *Neuroimage*, **61**(2), 397–406.
- Burnett, S., Bird, G., Moll, J., Frith, C., Blakemore, S.J. (2009). Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience*, **21**(9), 1736–50.
- 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 and Biobehavioral Reviews*, **35**(8), 1654–64.
- Carrington, S.J., Bailey, A.J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, **30**(8), 2313–35.
- Chen, G., Saad, Z.S., Britton, J.C., Pine, D.S., Cox, R.W. (2013). Linear mixed-effects modeling approach to fMRI group analysis. *Neuroimage*, **73**, 176–90.
- 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), 218–36.
- Corwin, J. (1994). On measuring discrimination and response bias: unequal numbers of targets and distractors and two classes of distractors. *Neuropsychology*, **8**(1), 110–7.
- Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, **29**(3), 162–73.
- Cox, R.W., Reynolds, R.C., Taylor, P.A. (2016). AFNI and clustering: false positive rates redux. *Brain Connectivity*, **7**(3), 152–71.
- Crone, E.A., Dahl, R.E. (2012). Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nature Reviews Neuroscience*, **13**, 636–50.

- Delgado, M.R., Locke, H.M., Stenger, V.A., Fiez, J.A. (2003). Dorsal striatum responses to reward and punishment: effects of valence and magnitude manipulations. *Cognitive, Affective & Behavioral Neuroscience*, **3**(1), 27–38.
- Dricu, M., Frühholz, S. (2016). Perceiving emotional expressions in others: activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. *Neuroscience and Biobehavioral Reviews*, **71**, 810–828.
- Eklund, A., Nichols, T.E., Knutsson, H.J.P.O.T.N.A.O.S. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, **113**(28), 7900–5.
- Ethofer, T., Anders, S., Wiethoff, S., et al. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *NeuroReport*, **17**(3), 249–53.
- Ethofer, T., Kreifelts, B., Wiethoff, S., et al. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, **21**(7), 1255–68.
- Forbes, E.E., Phillips, M.L., Silk, J.S., Ryan, N.D., Dahl, R.E. (2011). Neural systems of threat processing in adolescents: role of pubertal maturation and relation to measures of negative affect. *Developmental Neuropsychology*, **36**(4), 429–52.
- Frith, C.D., Frith, U. (2006). The neural basis of mentalizing. *Neuron*, **50**(4), 531–4.
- Fruhholz, S., Ceravolo, L., Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, **22**(5), 1107–17.
- Frühholz, S., Grandjean, D. (2013). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience and Biobehavioral Reviews*, **37**(10), 2847–55.
- Giedd, J.N., Blumenthal, J., Jeffries, N.O., et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*, **2**(10), 861–3.
- Grosbras, M.-H., Ross, P.D., Belin, P. (2018). Categorical emotion recognition from voice improves during childhood and adolescence. *Scientific Reports*, **8**(1), 14791.
- Gunther Moor, B., Op de Macks, Z.A., Güroğlu, B., Rombouts, S.A.R.B., Van der Molen, M.W., Crone, E.A. (2011). Neurodevelopmental changes of reading the mind in the eyes. *Social Cognitive and Affective Neuroscience*, **7**(1), 44–52.
- Guyer, A.E., Choate, V.R., Detloff, A., et al. (2012). Striatal functional alteration during incentive anticipation in pediatric anxiety disorders. *American Journal of Psychiatry*, **169**(2), 205–12.
- Hooker, C.I., Verosky, S.C., Germine, L.T., Knight, R.T., D'Esposito, M. (2008). Mentalizing about emotion and its relationship to empathy. *Social Cognitive and Affective Neuroscience*, **3**(3), 204–17.
- Iredale, J.M., Rushby, J.A., McDonald, S., Dimoska-Di Marco, A., Swift, J. (2013). Emotion in voice matters: neural correlates of emotional prosody perception. *International Journal of Psychophysiology*, **89**(3), 483–90.
- Johnson, M.H. (2000). Functional brain development in infants: elements of an interactive specialization framework. *Child Development*, **71**(1), 75–81.
- 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: Lewis, M., Haviland, J., editors. *Handbook of Emotions*. New York: Guilford, 220–35.
- Kaplan, J.T., Aziz-Zadeh, L., Uddin, L.Q., Iacoboni, M. (2008). The self across the senses: an fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective Neuroscience*, **3**(3), 218–23.
- Kilford, E.J., Garrett, E., Blakemore, S.J. (2016). The development of social cognition in adolescence: an integrated perspective. *Neuroscience and Biobehavioral Reviews*, **70**, 106–20.
- King, K.M., Littlefield, A.K., McCabe, C.J., Mills, K.L., Flournoy, J., Chassin, L. (2018). Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology. *Developmental Cognitive Neuroscience*, **33**, 54–72.
- Kirby, L.A., Robinson, J.L. (2017). Affective mapping: an activation likelihood estimation (ALE) meta-analysis. *Brain and Cognition*, **118**, 137–48.
- 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. *Hormones and Behavior*, **64**(2), 314–22.
- Kobayashi, C., Glover, G.H., Temple, E. (2007). Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. *Neuropsychologia*, **45**(7), 1522–32.
- Kotz, S.A., Kalberlah, C., Bahlmann, J., Friederici, A.D., Haynes, J.D. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping*, **34**(8), 1971–81.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.J.J.O.S.S. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13), 1–26.
- 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.
- Lee, T.-H., Perino, M.T., McElwain, N.L., Telzer, E.H. (2020). Perceiving facial affective ambiguity: a behavioral and neural comparison of adolescents and adults. *Emotion*, **20**(3), 501–6.
- Leppänen, J.M., Nelson, C.A. (2006). The development and neural bases of facial emotion recognition. In: Kail, R.V., editor. *Advances in Child Development and Behavior*, Vol. 34, Cambridge, Massachusetts, United States: Academic Press, JAI, 207–46.
- Li, W., Mai, X., Liu, C. (2014). The default mode network and social understanding of others: what do brain connectivity studies tell us. *Frontiers in Human Neuroscience*, **8**, 74.
- Liebenthal, E., Silbersweig, D.A., Stern, E. (2016). The language, tone and prosody of emotions: neural substrates and dynamics of spoken-word emotion perception. *Frontiers in Neuroscience*, **10**, 506.
- Luke, S.G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, **49**(4), 1494–502.
- Meyer, M.L. (2019). Social by default: characterizing the social functions of the resting brain. *Current Directions in Psychological Science*, **28**(4), 380–6.
- Mills, K.L., Lalonde, F., Clasen, L.S., Giedd, J.N., Blakemore, S.J. (2014). Developmental changes in the structure of the social brain in late childhood and adolescence. *Social Cognitive and Affective Neuroscience*, **9**(1), 123–31.
- Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*, 1st edn. Boca Raton, Florida, United States: Chapman and Hall/CRC.
- Mitchell, R.L.C., Phillips, L.H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, **70**, 1–10.
- Mitchell, R.L.C., Ross, E.D. (2013). Attitudinal prosody: what we know and directions for future study. *Neuroscience and Biobehavioral Reviews*, **37**(3), 471–9.
- Moore, W.E., 3rd, Pfeifer, J.H., Masten, C.L., Mazziotta, J.C., Iacoboni, M., Dapretto, M. (2012). Facing puberty: associations between pubertal development and neural responses to affective facial displays. *Social Cognitive and Affective Neuroscience*, **7**(1), 35–43.

- 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), 155–83.
- 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), 237–51.
- Morningstar, M., Nelson, E.E., Dirks, M.A. (2018b). Maturation of vocal emotion recognition: insights from the developmental and neuroimaging literature. *Neuroscience and Biobehavioral Reviews*, **90**, 221–30.
- Morningstar, M., Mattson, W.I., Venticinque, J., et al. (2019). Age-related differences in neural activation and functional connectivity during the processing of vocal prosody in adolescence. *Cognitive, Affective & Behavioral Neuroscience*, **19**(6), 1418–32.
- Nakamura, K., Kawashima, R., Ito, K., et al. (1999). Activation of the right inferior frontal cortex during assessment of facial emotion. *Journal of Neurophysiology*, **82**(3), 1610–4.
- 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**, 163–74.
- Nelson, E.E., Jarcho, J.M., Guyer, A.E. (2016). Social re-orientation and brain development: an expanded and updated view. *Developmental Cognitive Neuroscience*, **17**, 118–27.
- Overgaauw, S., van Duijvenvoorde, A.C.K., Gunther Moor, B., Crone, E.A. (2014). A longitudinal analysis of neural regions involved in reading the mind in the eyes. *Social Cognitive and Affective Neuroscience*, **10**(5), 619–27.
- Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, **9**(2), 60–8.
- Peelen, M.V., Atkinson, A.P., Andersson, F., Vuilleumier, P. (2007). Emotional modulation of body-selective visual areas. *Social Cognitive and Affective Neuroscience*, **2**(4), 274–83.
- Pfeifer, J.H., Masten, C.L., Moore, W.E., 3rd, et al. (2011). Entering adolescence: resistance to peer influence, risky behavior, and neural changes in emotion reactivity. *Neuron*, **69**(5), 1029–36.
- 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.
- R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rausch, J.R., Maxwell, S.E., Kelley, K. (2003). Analytic methods for questions pertaining to a randomized pretest, posttest, follow-up design. *Journal of Clinical Child and Adolescent Psychology*, **32**(3), 467–86.
- Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neuroscience and Biobehavioral Reviews*, **32**(1), 123–42.
- Ross, P., de Gelder, B., Crabbe, F., Grosbras, M.-H. (2019). Emotion modulation of the body-selective areas in the developing brain. *Developmental Cognitive Neuroscience*, **38**, 100660.
- Saxe, R., Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *Neuroimage*, **19**(4), 1835–42.
- Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, **43**(10), 1391–9.
- Scherer, K.R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Communication*, **40**(1–2), 227–56.
- Schirmer, A. (2018). Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing. *Social Cognitive and Affective Neuroscience*, **13**(1), 1–13.
- Schirmer, A., Kotz, S.A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, **10**(1), 24–30.
- Schriber, R.A., Guyer, A.E. (2016). Adolescent neurobiological susceptibility to social context. *Developmental Cognitive Neuroscience*, **19**, 1–18.
- Sebastian, C.L., Fontaine, N.M.G., Bird, G., et al. (2011). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, **7**(1), 53–63.
- Sergerie, K., Lepage, M., Armony, J.L. (2005). A face to remember: emotional expression modulates prefrontal activity during memory formation. *Neuroimage*, **24**(2), 580–5.
- Shamay-Tsoory, S.G., Aharon-Peretz, J., Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, **132**(3), 617–27.
- Snodgrass, J.G., Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology. General*, **117**(1), 34.
- Spear, L.P. (2000). The adolescent brain and age-related behavioral manifestations. *Neuroscience and Biobehavioral Reviews*, **24**(4), 417–63.
- Steinberg, L., Morris, A.S. (2001). Adolescent development. *Annual Review of Psychology*, **52**, 83–110.
- Tahmasebi, A.M., Artiges, E., Banaschewski, T., et al. (2012). Creating probabilistic maps of the face network in the adolescent brain: a multicentre functional MRI study. *Human Brain Mapping*, **33**(4), 938–57.
- Taylor, S.J., Barker, L.A., Heavey, L., McHale, S. (2015). The longitudinal development of social and executive functions in late adolescence and early adulthood. *Frontiers in Behavioral Neuroscience*, **9**, 252.
- Wang, A.T., Lee, S.S., Sigman, M., Dapretto, M. (2006). Developmental changes in the neural basis of interpreting communicative intent. *Social Cognitive and Affective Neuroscience*, **1**(2), 107–21.
- Wildgruber, D., Ackermann, H., Kreifelts, B., Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, **156**, 249–68.
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv:1308.5499*, 1–42. Available: <http://arxiv.org/pdf/1308.5499.pdf>.
- Yovel, G., Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, **17**(6), 263–71.
- Zaki, J., Weber, J., Bolger, N., Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences*, **106**(27), 11382–7.
- Zaki, J., Weber, J., Ochsner, K. (2012). Task-dependent neural bases of perceiving emotionally expressive targets. *Frontiers in Human Neuroscience*, **6**, 228.

Appendix

For many regions, happy voices were found to elicit greater activation relative to neutral, whereas sad/fearful/angry voices elicited lesser activation than neutral voices. This pattern was

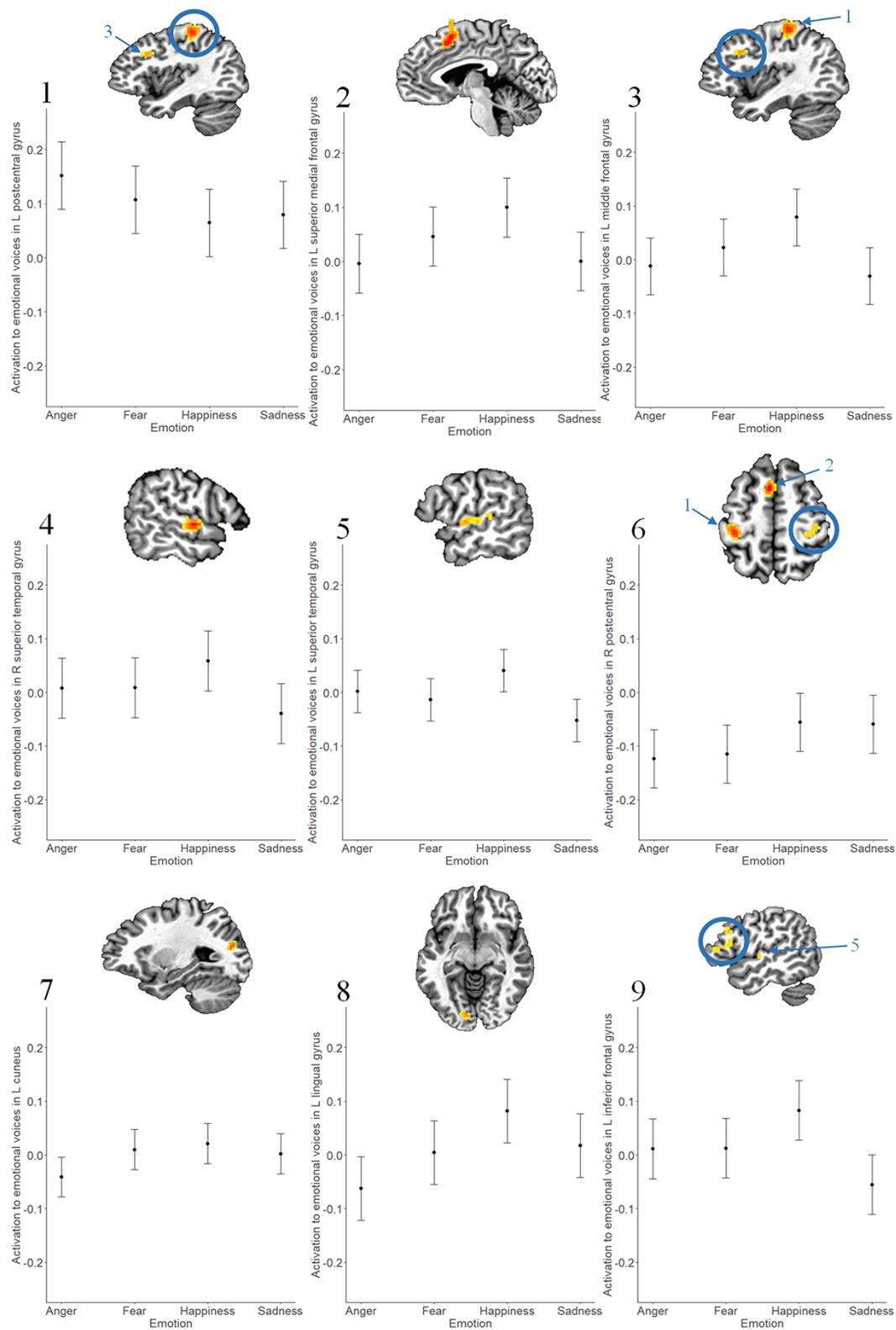


Fig. A. Effect of Emotion on neural activation to emotional voices.

Note. L = left, R = right. Clusters represent areas in which there was an effect of Emotion on activation to emotional voices (vs neutral voices). Brain images are rendered in the Talairach-Tournoux template space. Clusters were formed using 3dclustsim at $P < 0.001$ (corrected, with a cluster-size threshold of 26 voxels). Refer to Table 3 for a description of regions of activation. Error bars represent the standard error of the mean.

observed in the bilateral superior temporal gyrus (clusters 4 and 5), the dorsolateral prefrontal cortex at the midline of the brain (cluster 2), as well as the left inferior frontal gyrus (cluster 9), the left medial-lateral prefrontal cortex (cluster 3), the left inferior temporal gyrus (cluster 7) and the left lingual gyrus (cluster 8).

Significant *post hoc* comparisons between emotion types in each of these areas are illustrated in [Figure A](#). In addition, there was also an effect of emotion type on the bilateral postcentral gyrus (clusters 1 and 6), which likely reflected motor preparation for different response types.