ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202009.00054

Brain Mechanisms of Explicit and Implicit Emotional Prosody Processing: A Near-Infrared Imaging Study

Authors: Lei Zhen, Bi Rong, Mo Licheng, Yu Wenwen, Zhang Dandan, Zhang Dandan

Date: 2020-09-18T00:00:00+00:00

Abstract

Accurate recognition of emotional prosody information in speech is crucial for social interaction. This study employed functional near-infrared spectroscopy (fNIRS) to investigate cortical neural activity during the processing of angry, fearful, and happy prosody under explicit and implicit emotional processing conditions. The results indicated that the brain regions specifically engaged in processing angry, fearful, and happy prosody were the left frontopolar/orbitofrontal cortex, right supramarginal gyrus, and left inferior frontal gyrus, respectively, with the right supramarginal gyrus being concurrently modulated by both emotion and task demands. Additionally, the right middle temporal gyrus, inferior temporal gyrus, and temporal pole exhibited significantly greater activation during explicit emotional tasks compared to implicit tasks. The findings partially support the hierarchical model of emotional prosody processing while also challenging the model' s third-level assertion that "fine-grained processing of vocal emotional information in frontal regions requires participation of explicit emotional processing tasks."

Full Text

Brain Mechanisms of Explicit and Implicit Emotional Prosody Processing: An fNIRS Study

Lei Zhen[1], Bi Rong[2], Mo Licheng[2], Yu Wenwen[2], Zhang Dandan[1],[2]

[1] China Center for Behavioral Economics and Behavior Finance, Southwestern University of Finance and Economics, Chengdu 611130, China
[2] School of Psychology, Shenzhen University, Shenzhen 518060, China

Abstract
Accurate identification of emotional prosody in speech is crucial for social interaction. This study employed functional near-infrared spectroscopy to explore cortical neural activity during the processing of angry, fearful, and happy emotional prosodies under explicit and implicit emotional processing conditions. The results revealed that brain regions specifically processing angry, fearful, and happy prosodies were the left frontopolar/orbitofrontal cortex, right supramarginal gyrus, and left inferior frontal gyrus, respectively, with the right supramarginal gyrus being modulated by both emotion and task. Additionally, the right middle temporal gyrus, inferior temporal gyrus, and temporopolar area showed significantly stronger activation during the explicit emotional task compared to the implicit task. These findings partially support the hierarchical model of emotional prosody processing while challenging the model' s third level —that "fine-grained processing of emotional information in speech by frontal regions requires explicit emotional processing tasks."

Keywords: emotion; speech prosody; superior temporal gyrus; orbitofrontal cortex; inferior frontal gyrus; supramarginal gyrus

Accurately decoding emotional information is not only essential for human survival but also helps us perceive others' emotional states and infer their intentions during social interactions. Common carriers for conveying emotions in daily communication include facial expressions, body postures, and vocal cues. While the neural mechanisms underlying the processing of facial and bodily emotional information via visual channels have been relatively well characterized (see reviews by Calvo & Nummenmaa, 2016; Enea & Iancu, 2016; Hinojosa, Mercado, & Carretié, 2015), the mechanisms for decoding emotional information from speech remain less understood (Liebenthal, Silbersweig, & Stern, 2016). Emotional information in speech can be expressed through both semantic content and prosodic features. This study aimed to investigate the neural mechanisms of emotional prosody processing. Emotional prosody refers to the expression and communication of emotions through dynamic variations in acoustic cues such as pitch, duration, intensity, stress, and intonation, independent of lexical and grammatical structure (Brük, Kreifelts, & Wildgruber, 2011). These prosodic cues not only help us comprehend speakers' semantic and emotional states but also enable us to infer their intentions when prosodic and semantic information conflict (Ben-David, Multani, Shakuf, Rudzicz, & van Lieshout, 2016). For example, when someone says "I' m so lucky" with an angry prosody, we perceive the speaker as expressing anger. Thus, processing emotional prosody in speech is closely related to our daily lives and social interactions (Brük et al., 2011; Früholz, Trost, & Kotz, 2016). Elucidating the neural mechanisms of emotional prosody processing not only advances our understanding of the brain's emotional speech processing but also facilitates early diagnosis and rehabilitation assessment of patients with social dysfunction (e.g., autism, schizophrenia, depression) (Knight & Baune, 2019; Lin, Ding, & Zhang, 2018). Furthermore, revealing the brain' s decoding rules for emotional prosody forms the basis for synthesizing emotional speech intonation in artificial systems, which will greatly promote the

development of artificial intelligence and human-computer interaction interfaces (Mitchell & Xu, 2015).

Early models of emotional prosody processing (Ross, 1981) proposed that emotional prosody is processed exclusively by the right hemisphere, with the right superior temporal cortex (STC)[2] corresponding to Wernicke' s area in the left hemisphere responsible for perceiving and comprehending emotional prosody, and the right dorsolateral prefrontal cortex corresponding to Broca' s area in the left hemisphere responsible for producing emotional speech prosody. As research progressed, scholars gradually discovered that other cortical and subcortical structures also participate in emotional prosody processing, leading to the development of a hierarchical model (Brük et al., 2011; Ethofer et al., 2006; Schirmer & Kotz, 2006; Witteman, Van Heuven, & Schiller, 2012). This model posits three levels of emotional prosody processing: (1) the mid-portion of the sound-sensitive STC (mid-STC) receives neural signals from primary auditory cortex and extracts acoustic parameters from speech; (2) the posterior portion of the right STC (p-STC) identifies emotional information in speech through multimodal integration; and (3) frontal regions, represented by bilateral inferior frontal gyrus (IFG) and orbitofrontal cortex (OFC), evaluate and perform fine-grained processing of emotional information in speech. The hierarchical model suggests that the transmission of auditory information from primary auditory cortex to mid-STC represents bottom-up, stimulus-driven processing, whereas integration of speech information and fine-grained emotional processing in p-STC and frontal regions depend on attention and explicit emotional evaluation (Brük et al., 2011). Furthermore, Früholz et al. (2016) conducted a meta-analysis of studies on the neural mechanisms of emotional sounds (including non-human sounds, human nonverbal vocalizations, prosodic speech, and music) and concluded that emotional sound processing relies on a widely distributed neural network with core regions including the amygdala (Früholz, Hofstetter, Cristinzio, Saj, Seeck, & Vuilleumier, 2015), primary and secondary auditory cortex, right STC (Früholz & Grandjean, 2013a), IFG (Früholz & Grandjean, 2013b), OFC (Kotz, Kalberlah, Bahlmann, Friederici, & Haynes, 2013), and insula (Mothes-Lasch, Mentzel, Miltner, & Straube, 2011).

We believe three issues regarding the neural mechanisms of emotional prosody processing require resolution. First, the cerebral representation of different emotion categories in speech prosody. Previous research has shown that processing basic emotions such as anger, disgust, fear, happiness, and sadness relies on partially non-overlapping brain regions, indicating relative specificity in how the brain processes different emotion categories (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). However, this conclusion is primarily based on evidence from visual modalities (e.g., facial expressions, body postures, emotional word processing). In the auditory domain, particularly for emotional prosody processing, only two relevant studies have employed machine learning algorithms in a data-driven approach to classify fMRI activation patterns while participants listened to different emotional prosodies (Ethofer, Van De Ville, Scherer, & Vuilleumier, 2009a; Kotz et al., 2013). These studies found that brain re-

gions contributing to multi-class classification (distinguishing different emotion categories) included the superior temporal gyrus (STG), superior temporal sulcus (STS), middle temporal gyrus (MTG), IFG, middle frontal gyrus (MFG), and anterior insula, suggesting these regions exhibit some discriminability (i.e., specificity) for processing different emotional prosodies. Additionally, Kotz et al. (2013) compared brain regions activated by multiple emotional prosodies versus neutral prosody and found that each emotion activated distinct regions; for example, angry prosody activated bilateral MFG more strongly than neutral prosody, while happy prosody activated left IFG more strongly than neutral prosody. However, beyond these three studies, most existing research on emotional prosody processing has focused only on differences between a specific emotional prosody and neutral prosody, rarely examining the specific processing of different emotion categories (Früholz & Grandjean, 2013a). We argue that comparing a specific emotion to a neutral condition alone cannot identify "specific" brain regions for emotional prosody processing. This study investigated the specific brain regions processing angry, fearful, and happy prosodies —three emotions with high discriminability (Liu & Pell, 2012) and representativeness. Specifically, we conducted pairwise comparisons among the three emotional prosody conditions to identify brain regions involved exclusively in processing a single emotion category. Based on relevant literature (Früholz & Grandjean, 2013b), we expected to find brain regions in frontal areas such as IFG that could discriminate among the three emotional prosodies.

Second, the similarities and differences in brain activation between "emotion-relevant" and "emotion-irrelevant" tasks. Although some studies have compared brain regions involved in emotional prosody processing under explicit and implicit tasks, contradictory results remain. For instance, some studies found that implicit tasks activated p-STG while explicit tasks activated mid-STC (Früholz, Ceravolo, & Grandjean, 2012), contrary to the hierarchical model' s predictions. Additionally, many studies reported stronger activation in emotional brain regions such as the amygdala and insula during explicit versus implicit tasks (Früholz et al., 2012), while others observed main effects of emotion (i.e., stronger neural activity for angry prosody than neutral prosody) in these regions under both explicit and implicit tasks (Bach et al., 2008; Ethofer et al., 2009b; Quadfieg, Mohr, Mentzel, Miltner, & Straube, 2008). Importantly, findings in frontal regions have also been inconsistent: some studies reported activation of bilateral or unilateral IFG during both explicit and implicit emotional tasks (Früholz et al., 2012; Steber, König, Stephan, & Rossi, 2020), while others observed stronger activation in OFG (Ethofer et al., 2009b; Quadfieg et al., 2008) and IFG (Bach et al., 2008; Beaucousin et al., 2011) during explicit tasks. We believe these contradictions may first arise from inappropriate task designs. Most previous studies used "emotion discrimination" as the explicit task and "gender discrimination" as the implicit task. Compared to emotion discrimination, gender discrimination is simpler, so participants may have involuntarily engaged in emotional and semantic processing during the gender discrimination task, introducing confounding variables. This study replaced "gender discrimi-

nation" with "identity discrimination" —all speech materials were female voices, and participants discriminated between different female speakers, increasing the difficulty of the implicit task and reducing the likelihood of emotional processing. Second, most previous studies used words as stimuli (e.g., Bach et al., 2008; Ethofer et al., 2009b; Früholz et al., 2012; Quadfieg et al., 2008; Steber et al., 2020). Given that emotional prosody is a long-term characteristic of speech and single-word presentation times are very limited (550–750 ms, mostly two-syllable words), we believe longer speech units (e.g., sentences) should be used in emotional prosody research. Additionally, some studies used semantic materials (Beaucousin et al., 2011; Ethofer et al., 2009b; Mitchell, 2007; Quadfieg et al., 2008), while others used non-semantic materials (Bach et al., 2008; Früholz et al., 2012; Steber et al., 2020), which may also interfere with results. Therefore, this study used non-semantic "pseudo-sentences." By increasing the cognitive load of the implicit task, extending the presentation time of emotional prosody materials, and eliminating semantic processing interference, this study aimed to more accurately reveal differences in brain mechanisms between explicit and implicit emotional prosody processing. Given the numerous contradictory findings in existing literature, we could not propose specific hypotheses for this issue (exploratory question).

Third, further accumulation of experimental evidence based on noise-free brain imaging technology is needed. Most previous studies on the neural mechanisms of emotional prosody processing have used fMRI technology (Früholz et al., 2016). While fMRI is a mainstream neuroimaging technique that provides accurate spatial localization of whole-brain neural activity, it generates high-intensity noise during scanning due to gradient coil positioning, which can affect tasks requiring attention to acoustic parameter details such as emotional prosody processing (Dieler, Tupak, & Fallgatter, 2012). Therefore, we suggest that fMRI findings based on auditory perception channels should ideally be validated by other brain imaging techniques. Functional near-infrared spectroscopy (fNIRS) features noise-free operation and relative insensitivity to head movement, making it highly suitable for speech processing research. This study aimed to obtain brain regions involved in emotional prosody processing consistent with previous fMRI studies using fNIRS technology.

2.1 Participants

Sample size was estimated using G*Power 3.1.7 software. Referring to effect sizes from similar studies (partial $^2 > 0.10$) (Bach et al., 2008; Früholz et al., 2012), a statistical power of 0.95 could be achieved with 16 participants. Additionally, referring to our research group' s previous study sample size (n = 22) (Zhang, Zhou, & Yuan, 2018), this study recruited 25 university students (13 female; age: 20.42 ± 2.13 years). All participants had normal hearing and were right-handed. Exclusion criteria were: (1) history of psychiatric disorders; (2) history of head trauma or epilepsy; (3) severe physical illness; (4) alcohol or drug dependence. Participants were informed about the experimental equipment and

procedures, provided written informed consent, and received monetary compensation after the experiment. The experimental protocol was approved by the Shenzhen University Ethics Committee.

2.2 Experimental Materials and Procedure

Emotional prosody materials were selected from the Chinese Speech Emotion Database (Liu & Pell, 2012). This database contains emotional prosody speech in the form of grammatically correct but meaningless "pseudo-sentences" : each sentence follows subject-verb-object structure but uses meaningless "pseudo-words" for all components, thereby eliminating semantic content. Each sentence lasted approximately 1–2 s. This study selected materials with angry, fearful, happy, and neutral prosodies (30 pseudo-sentences per emotion condition). For each emotion, 5–9 sentences were combined into 10-s speech segments. Each 10-s emotional speech segment consisted of 5–8 emotional sentences (of the same emotion) and 1–4 neutral sentences (filler stimuli), with no intervals between pseudo-sentences. One hundred twenty pseudo-sentences were used to create five 10-s segments for each of angry, fearful, and happy prosodies, with no pseudo-sentence repeated across segments. Each 10-s speech segment contained sentences spoken by two female speakers, with unequal numbers of sentences corresponding to the two identities across the 15 segments.

This experiment used a 3 (emotion: fear, anger, happiness) × 2 (task: explicit, implicit) within-subjects design. Speech stimuli were presented through two speakers (EDIFIER-R26T, Dongguan, China) placed approximately 50 cm in front of the participant' s left and right ears. Speech materials were presented at 60–70 dB sound intensity, with an average background noise level of 30 dB (when no speech was presented).

The experiment consisted of two tasks, with task order counterbalanced across participants. The **explicit emotional processing task** required participants to count emotional and neutral sentences in each 10-s speech segment and respond after each segment: "Are there 3 or more emotional sentences than neutral sentences?" (Yes/No answers each accounted for 50%). This task included three blocks (fear, anger, happiness), with block order randomized across participants. Each block contained 10 ten-second speech segments (each speech material was presented twice), presented in random order, with a silent interval of 15 s minus response time between segments. The **implicit emotional processing task** required participants to count sentences spoken by the two female speakers in each 10-s segment and respond: "Do the two female speakers differ by more than 2 sentences?" (Yes/No answers each accounted for 50%). This task also included three blocks, each containing 10 ten-second speech segments. In both tasks, each block lasted approximately 4 min, with self-paced rest periods between blocks (2–10 min). To control for confounding effects of speech and semantic variables, the 30 ten-second speech segments used in both tasks were identical.

2.3 fNIRS Data Recording

A multi-channel fNIRS system recorded brain activity in continuous wave mode (NirScan, HuiChuang, China). Optical probe positioning used a NIRS-EEG compatible cap based on the international 10/20 system (EASYCAP, Herrsching, Germany). Based on existing literature, this study observed bilateral frontal and temporal lobes. We used 13 emitters and 15 detectors to form 37 effective observation channels [Figure 1: see original paper], with an average emitter-detector distance of 3.2 cm (range: 2.8–3.6 cm). No detector signal saturation occurred during recording. We defined the midpoint of each channel as the primary brain region detected by that channel and used this point as the center for channel localization. First, the NFRI toolbox (http://brain.job.affrc.go.jp/tools/) was used to calculate MNI coordinates for each channel, which were then mapped to brain regions using the adult Brodmann Talairach brain template (Lancaster et al., 2000). Channel coordinates and localization information are shown in Table 1 .

2.4 fNIRS Data Analysis

Data preprocessing was performed using NirSpark software (HuiChuang, China). (1) Motion artifacts in raw optical density data were corrected using a wavelet-based method (Molavi & Dumont, 2012); (2) Data were band-pass filtered at 0.01–0.20 Hz; (3) Filtered optical density data were converted to concentration changes of HbO and HbR (Δ[HbO] and Δ[HbR]) based on the modified Beer-Lambert law. In this study, Δ[HbO] had higher signal-to-noise ratio and was more sensitive to cerebral blood flow changes than Δ[HbR] (Tong, Hocke, & Frederick, 2011; Zhang, Chen, Hou, & Wu, 2019), so subsequent statistical analyses used Δ[HbO] data.

A general linear model (GLM) was used to estimate task-related β values for different conditions, with β values serving as indicators of brain region activation. A block design was employed, using the 10-s duration of each speech segment as the block length for convolution with the hemodynamic response function (HRF). Statistical analyses were conducted using SPSS 20.0 software (IBM, Somers, USA). The significance level was set at 0.05. Descriptive statistics are reported as mean $\pm$ standard deviation. A 3 (emotion: fear, anger, happiness) $\times$ 2 (task: explicit, implicit) repeated measures ANOVA was performed on β values for each channel, with Greenhouse-Geisser correction for sphericity and Bonferroni correction for post-hoc multiple comparisons. Finally, false discovery rate (FDR) correction was applied to p-values across channels to further reduce false positive rates.

Reaction time (628.35 $\pm$ 151.11 ms) and accuracy (91.42% $\pm$ 6.30%) did not differ significantly across experimental conditions ($F < 1$).

Brain activation results are as follows. First, we found main effects of emotion in channels 3, 9, and 32 , [Figure 2: see original paper], and [Figure 3: see original paper]A. Post-hoc comparisons revealed that the left frontopolar and

orbitofrontal cortex (frontopolar/orbitofrontal area; i.e., anterior portions of the middle and superior frontal gyri, channel 3) was most sensitive to angry speech ($F(2,48) = 12.51$, $p < 0.001$, partial $^2 = 0.343$; corrected $p = 0.001$; anger vs. fear $p = 0.007$; anger vs. happiness $p = 0.001$). The left pars triangularis of the inferior frontal gyrus (part of Broca' s area; channel 9) was most sensitive to happy speech ($F(2,48) = 24.26$, $p < 0.001$, partial $^2 = 0.502$; corrected $p < 0.001$; happiness vs. anger/fear ps < 0.001). The right supramarginal gyrus (SMG; channel 32) was most sensitive to fearful speech ($F(2,48) = 12.53$, $p < 0.001$, partial $^2 = 0.343$; corrected $p = 0.001$; fear vs. anger $p = 0.001$; fear vs. happiness $p = 0.003$).

Second, we found main effects of task in channels 27–29 , [Figure 2: see original paper], and [Figure 3: see original paper]B. In three right temporal channels, explicit tasks elicited stronger cortical activity than implicit tasks. Channel 27 corresponded to the temporopolar area ($F(1,24) = 11.63$, $p = 0.002$, partial $^2 = 0.325$; corrected $p = 0.004$), channel 28 to the superior temporal gyrus (STG; $F(1,24) = 26.10$, $p < 0.001$, partial $^2 = 0.521$; corrected $p < 0.001$), and channel 29 to the middle temporal gyrus (MTG; $F(1,24) = 15.81$, $p = 0.001$, partial $^2 = 0.397$; corrected $p = 0.003$).

Finally, we found an interaction between task type and emotion in channel 32 ($F(2,48) = 14.24$, $p < 0.001$, partial $^2 = 0.372$; corrected $p = 0.002$; [Figure 2: see original paper] and [Figure 3: see original paper]C), which corresponded to the right SMG. Simple effects analysis showed that in the explicit task, the right SMG was more sensitive to fearful (β: 0.63 ± 0.43) than to angry (0.06 ± 0.44) and happy prosodies (0.10 ± 0.34) ($F(2,23) = 102.27$, $p < 0.001$, partial $^2 = 0.898$), whereas this emotion effect was not significant in the implicit task ($F < 1$; fear/anger/happiness β values = 0.10 ± 0.44 / 0.14 ± 0.66 / 0.12 ± 0.55).

Table 2. Main Effects of Emotion

Brain Region	Anger β	Fear β	Happiness β	p (FDR-corrected)
L Frontopolar/orbitofrontal area	0.21 ± 0.20	0.12 ± 0.23	0.06 ± 0.20	$< .001$
L pars triangularis/Broca' s area	0.10 ± 0.16	0.10 ± 0.15	0.21 ± 0.15	$< .001$
R Supramarginal gyrus	0.10 ± 0.56	0.36 ± 0.51	0.11 ± 0.45	$< .001$

The p-values were corrected using FDR method across fNIRS channels.

Table 3. Main Effects of Task

Brain Region	Implicit β	Explicit β	p (FDR-corrected)
R Temporopolar area	0.04 ± 0.36	0.32 ± 0.42	0.004
R Superior temporal gyrus	0.05 ± 0.45	0.37 ± 0.43	< 0.001
R Middle temporal gyrus	-0.03 ± 0.49	0.34 ± 0.53	0.003

The p-values were corrected using FDR method across fNIRS channels.

Figure 2. Activation in different brain regions across emotion and task conditions (only channels showing significant effects are displayed). Error bars represent standard error of the mean.

Figure 3. Imaging maps of brain activation.

This study used fNIRS technology to investigate cortical neural activity during emotional prosody processing under explicit and implicit emotional tasks. The results showed that explicit emotional tasks elicited stronger activation in right temporal STG, MTG, and temporopolar area than implicit tasks, indicating that the right temporal lobe plays an important role in task-relevant emotional prosody perception. We also found that discriminative decoding of different emotion categories depended on frontal IFG and OFC, and these regions were activated under both explicit and implicit emotional tasks. In contrast, the parietal SMG region discriminated fearful prosody only during explicit tasks.

The main effect of emotion indicated that the brain can distinguish different emotional prosody categories under both task-relevant and task-irrelevant conditions. Compared to fearful and happy prosodies, angry prosody more strongly activated the left frontopolar and OFC region. This finding is highly similar to previous results from emotional prosody discrimination tasks (Kotz et al., 2013) and passive listening tasks (Zhang et al., 2018). The OFC not only plays an important role in anger processing (Lindquist et al., 2012) but is also responsible for conflict resolution and inhibiting inappropriate behaviors (e.g., aggression) (Beyer, Münte, Göttlich, & Krämer, 2015). Patients with OFC damage show increased aggression and significant deficits in subjective emotional state evaluation, emotional information integration, and emotional prosody recognition (Fox et al., 2018; Herpertz et al., 2017; Paulmann, Seifert, & Kotz, 2010).

Similar to the OFC findings for angry prosody, we also found that the left frontal IFG could specifically discriminate happy prosody from fearful and angry prosodies under both explicit and implicit task conditions, consistent with previous findings in adults (Zhang et al., 2018) and newborns (Zhang et al., 2019) during passive listening tasks. Specifically, our result in channel 9 localized to the pars triangularis of IFG, a region critical for semantic comprehension and integration of semantic-emotional prosody and other linguistic information (Goucha & Friederici, 2015; Kirby & Robinson, 2017; Schirmer & Kotz, 2006).

A study on sarcasm comprehension found that left IFG plays an important role in integrating context, semantics, and prosody, showing increased activation when positive semantics were paired with negative emotional prosody (i.e., sarcasm) (Matsui et al., 2016). In another study on auditory emotion and social judgment, left IFG was found to be involved in both social trait evaluation (trustworthiness) and speaker happiness evaluation (Hensel, Bzdok, Müller, Zilles, & Eickhoff, 2015). Two studies by Kotz et al. found that happy prosody activated left IFG compared to neutral prosody (Kotz et al., 2003; 2013), while Johnstone et al. observed enhanced activation in right IFG for happy prosody (Johnstone, Van Reekum, Oakes, & Davidson, 2006). The inconsistent hemispheric lateralization of IFG activation may be due to differences in experimental settings or emotional materials: Kotz et al. (2003; 2013) used speech prosody only, whereas Johnstone et al. (2006) required participants to view facial expressions that were emotionally congruent or incongruent with the speech prosody. Our emotional materials were similar to those in Kotz et al. (2003; 2013), and we found that happy prosody specifically activated left IFG. By conducting pairwise comparisons among emotional conditions, we identified "specific brain regions" for angry and happy emotional prosodies, a result not obtained in previous studies that only compared a specific emotion to a neutral condition.

According to the hierarchical model of emotional prosody processing, the third level involves fine-grained processing of emotional information in speech by frontal regions such as OFC and IFG, which depends on attention or explicit emotional evaluation (Brük et al., 2011). Our findings in OFC and IFG contradict this conclusion. Reviewing previous studies that used explicit and implicit emotional tasks to investigate emotional prosody processing, we found that studies observing stronger activation in OFC/IFG during explicit tasks mostly used semantic words (Ethofer et al., 2009b; Quadflieg et al., 2008) or sentences (Beaucousin et al., 2011) as experimental materials, whereas studies observing clear activation in frontal regions during both explicit and implicit tasks often used non-semantic emotional prosody materials (e.g., Früholz et al., 2012; Steber et al., 2020). We therefore infer that semantic processing may interfere with or influence emotional prosody processing. The only exceptional study we identified is Bach et al. (2008), who used pseudo-words as speech materials but still observed stronger activation in left IFG during explicit versus implicit conditions. However, that study used gender discrimination as the implicit task, which, as noted in the introduction, is much easier than emotion discrimination (accuracy and reaction time differed significantly between the two tasks), potentially affecting the results. Additionally, Bach et al. (2008) recruited only 16 participants, a relatively small sample size. This study used "identity discrimination" as the implicit task, matched task difficulty between implicit and explicit conditions, and employed non-semantic pseudo-sentences as experimental materials. Based on these methodological improvements, we found that frontal OFC/IFG regions could discriminate different emotional prosody categories under both explicit and implicit tasks, allowing us to confidently challenge the third-level processing theory of the hierarchical model.

Our results also showed that the right SMG region processes fearful prosody with relative specificity, an unexpected finding. SMG is involved in language perception and processing, and damage to this region causes sensory aphasia. SMG also participates in body posture processing and is considered part of the mirror neuron system (Carlson, 2012). The right SMG is specifically associated with speech prosody processing (Hartwigsen, Baumgaertner, Price, Koehnke, Ulmer, & Siebner, 2010) and emotion perception (Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, 2000; Aryani, Hsu, & Jacobs, 2018). Köchel et al. suggested that SMG is closely related to attention and alertness functions (Köchel, Schöngassner, & Schienle, 2013). Their study using fearful, disgusted, and neutral nonverbal sounds found that processing fearful sounds (e.g., screams of fear or pain) significantly increased activation in right STG and bilateral SMG. Additionally, Patel et al. (2018) found that impaired emotional prosody expression is associated with SMG damage. Most notably, SMG was the only brain region in this study that was modulated by both emotion and task: right SMG showed specific processing of fearful prosody only during explicit emotional tasks. On one hand, we did not observe right SMG sensitivity to fearful prosody in our previous adult passive listening study (Zhang, Zhou, & Yuan, 2018), suggesting that processing fearful emotions in this region requires substantial task-relevant emotional engagement. On the other hand, the specific processing of fearful prosody in right SMG is consistent with our group's previous findings in newborns (Zhang, Zhou, Hou, Cui, & Zhou, 2017; Zhang et al., 2019), indicating that the human brain can rapidly respond to fearful emotional prosody at birth, independent of emotional task engagement. Combining our findings in adults and newborns regarding fearful emotional prosody, we propose that this may reflect an evolutionarily developed innate processing bias for fearful emotions —automatic brain responses to fearful emotions are present at birth, but with development, processing of fearful prosody (unlike fearful faces) requires active explicit emotional task engagement.

The main effect of task in this study showed that right MTG, STG, and temporopolar area were significantly more activated during explicit than implicit emotional tasks, a result fully consistent with the hierarchical model of emotional prosody processing (Bach et al., 2008; Brük et al., 2011; Ethofer et al., 2006; Schirmer & Kotz, 2006; Witteman et al., 2012) and with the right-lateralization theory of emotional speech prosody processing (see review by Belyk & Brown, 2014). The right STC is the primary structure of the "emotional voice area" (Ethofer et al., 2012; Liebenthal et al., 2016) and a key region for decoding emotional speech (see review by Früholz & Grandjean, 2013a). Lower-level structures of STC (primary auditory cortex and mid-STC) parse auditory features of emotional sounds (including speech, human nonverbal vocalizations, and natural sounds), while higher-level structures integrate these parsed acoustic features to construct perception of emotional speech (Früholz et al., 2016; Schirmer & Kotz, 2006). Previous studies have shown that STC is more strongly activated by emotional than neutral speech (Bach et al., 2008; Brück et al., 2011; Ethofer et al., 2009b; Früholz et al., 2012; Kotz et al., 2003; Mothes-Lasch et

al., 2011; Witteman et al., 2012). This study found a clear right-lateralization advantage of STC for explicit emotional prosody processing, consistent with existing literature (Früholz et al., 2016; Kotz et al., 2013) and indicating that the right hemisphere is sensitive to slow-varying signals and suprasegmental features in speech (Witteman et al., 2012). Given that right STG is a classic pitch-processing region (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002), our observed explicit task effect may also reflect active pitch discrimination processing in right STG during emotion discrimination tasks (angry, fearful, and happy prosodies all have higher pitch than neutral prosody; see acoustic material information in Liu & Pell, 2012). The right-lateralized temporal advantage for explicit emotional prosody processing found in this study may aid clinical diagnosis of brain disorders, such as identifying epileptic foci and assessing brain functional plasticity after epilepsy and other organic brain lesions (Alba-Ferrara, Kochen, & Hausmann, 2018).

In summary, this study identified specific brain regions for processing angry, fearful, and happy prosodies in left OFC, right SMG, and left IFG, respectively. We also demonstrated the important role of right STC (including MTG, STG, etc.) and right SMG in explicit emotional prosody tasks. Our findings partially support the hierarchical model of emotional prosody processing while challenging the model' s third level—that "fine-grained processing of emotional information in speech by frontal regions requires explicit emotional task engagement." Additionally, due to inherent limitations of fNIRS technology, this study could not detect deeper brain regions closely related to emotional prosody processing (e.g., superior temporal sulcus and amygdala). We suggest that future research in this field should combine multiple imaging techniques (fMRI, fNIRS, magnetoencephalography) with brain lesion studies and transcranial magnetic stimulation (particularly newly developed deep transcranial magnetic stimulation).

References

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *The Journal of Neuroscience, 20*(7), 2683–2690.

Alba-Ferrara, L., Kochen, S., & Hausmann, M. (2018). Emotional prosody processing in epilepsy: Some insights on brain reorganization. *Frontiers in Human Neuroscience, 12*, 92.

Aryani, A., Hsu, C. T., & Jacobs, A. M. (2018). The sound of words evokes affective brain responses. *Brain Sciences, 8*(6), 94.

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.

Beaucousin, V., Zago, L., Herve, P. Y., Strelnikov, K., Crivello, F., Mazoyer,

B., & Tzourio-Mazoyer, N. (2011). Sex-dependent modulation of activity in the neural networks engaged during emotional speech comprehension. *Brain Research, 1390*, 108–117.

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., & van Lieshout, P. H. (2016). Prosody and semantics are separate but not separable channels in the perception of emotional speech: test for rating of emotions in speech. *Journal of Speech Language and Hearing Research, 59*(1), 72–89.

Beyer, F., Munte, T. F., Göttlich, M., & Krämer, U. M. (2014). Orbitofrontal cortex reactivity to angry facial expression in a social interaction correlates with aggressive behavior. *Cerebral Cortex, 25*(9), 3057–3063.

Belyk, M., & Brown, S. (2014). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience, 9*, 1395–1403.

Brük, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews, 8*, 383–403.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion, 30*, 1081–1106.

Dieler, A. C., Tupak, S. V., & Fallgatter, A. J. (2012). Functional near-infrared spectroscopy for the assessment of speech related tasks. *Brain and Language, 121*(2), 90–109.

Enea, V., & Iancu, S. (2016). Processing emotional body expressions: state-of-the-art. *Social Neuroscience, 11*(5), 495–506.

Ethofer, T., Bretscher, 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. *Cerebral Cortex, 22*, 191–200.

Ethofer, T., et al. (2006). Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage, 30*, 580–587.

Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009a). Decoding of emotional information in voice-sensitive cortices. *Current Biology, 19*(12), 1028-1033.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009b). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience, 21*, 1255–1268.

Fox, K. C. R., Yih, J., Raccah, O., Pendekanti, S. L., Limbach, L. E., Maydan, D. D., & Parvizi, J. (2018). Changes in subjective experience elicited by direct stimulation of the human orbitofrontal cortex. *Neurology, 91*(16), e1519–e1527.

Früholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions - Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews, 68*, 96–110.

Früholz, S., & Grandjean, D. (2013a). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews, 37*, 24–35.

Früholz, S., & Grandjean, D. (2013b). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience and Biobehavioral Reviews, 37*(10), 2847–2855.

Früholz, S., Hofstetter, C., Cristinzio, C., Saj, A., Seeck, M., & Vuilleumier, P. (2015). Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. *Proceedings of the National Academy of Sciences of the United States of America, 112*(5), 1583–1588.

Früholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex, 22*, 1107–1117.

Goucha, T., & Friederici, A. D. (2015). The language skeleton after dissecting meaning: A functional segregation within Broca' s Area. *Neuroimage, 114*, 294–302.

Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., & Siebner, H. R. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proceedings of the National Academy of Sciences of the United States of America, 107*(38), 17495–17500.

Hensel, L., Bzdok, D., Müller, V. I., Zilles, K., & Eickhoff, S. B. (2015). Neural correlates of explicit social judgments on vocal stimuli. *Cerebral Cortex, 25*(5), 1152–1162.

Herpertz, S. C., Nagy, K., Ueltzhöffer, K., Schmitt, R., Mancke, F., Schmahl, C., & Bertsch, K. (2017). Brain mechanisms underlying reactive aggression in borderline personality disorder-sex matters. *Biological Psychiatry, 82*(4), 257–266.

Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. *Neuroscience and Biobehavioral Reviews, 55*, 498–509.

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. *Social Cognitive and Affective Neuroscience, 1*(3), 242–249.

Kirby, L. A. J., & Robinson, J. L. (2017). Affective mapping: An activation likelihood estimation (ALE) meta-analysis. *Brain and Cognition, 118*, 137–148.

Knight, M. J., & Baune, B. T. (2019). Social cognitive abilities predict psychosocial dysfunction in major depressive disorder. *Depression and Anxiety, 36*(1),

54–62.

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*, 1971–1981.

Kotz, S. A., et al. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain and Language, 86*, 366–376.

Köchel, A., Schöngassner, F., & Schienle, A. (2013). Cortical activation during auditory elicitation of fear and disgust: a near-infrared spectroscopy (NIRS) study. *Neuroscience Letters, 9*(549), 197–200.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., ···Fox, P. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping, 10*(3), 120–131.

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 Aging Neuroscience, 10*, 506.

Lin, Y., Ding, H., & Zhang, Y. (2018). Emotional prosody processing in schizophrenic patients: A selective review and meta-analysis. *Journal of Clinical Medicine, 7*(10), 363.

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: a validated database of Chinese vocal emotional stimuli. *Behavior Research Methods, 44*, 1042–1051.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences, 35*, 121–143.

Matsui, T., Nakamura, T., Utsumi, A., Sasaki, A. T., Koike, T., Yoshida, Y., & Harada, T., et al. (2016). The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence. *Neuropsychologia, 87*, 74–84.

Mitchell, R. L., & Xu, Y. (2015). What is the value of embedding artificial emotional prosody in human-computer interactions? Implications for theory and design in psychological science. *Frontiers in Psychology, 6*, 1750.

Mitchell, R. L. (2007). fMRI delineation of working memory for emotional prosody in the brain: commonalities with the lexico-semantic emotion network. *Neuroimage, 36*(3), 1015–1025.

Mothes-Lasch, M., Mentzel, H. J., Miltner, W. H. R., & Straube, T. (2011). Visual attention modulates brain activation to angry voices. *Journal of Neuroscience, 31*, 9594–9598.

Ross, E. D. (1981). The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology, 38*(9), 561–569.

Patel, S., Oishi, K., Wright, A., Sutherland-Foggio, H., Saxena, S., Sheppard, S. M., & Hillis, A. E. (2018). Right hemisphere regions critical for expression of emotion through prosody. *Frontiers in neurology, 9*, 224.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36*, 767–776.

Paulmann, S., Seifert, S., & Kotz, S. A. (2010). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Social Neuroscience, 5*(1), 59–75.

Quadflieg, S., Mohr, A., Mentzel, H. J., Miltner, W. H., & Straube, T. (2008). Modulation of the neural network involved in the processing of anger prosody: the role of task-relevance and social phobia. *Biological Psychology, 78*, 129–137.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*, 24–30.

Steber, S., König, N., Stephan, F., & Rossi, S. (2020). Uncovering electrophysiological and vascular signatures of implicit emotional prosody. *Scientific Reports, 10*(1), 5807.

Tong, Y., Hocke, L. M., & Frederick, B. deB., (2011). Isolating the sources of widespread physiological fluctuations in functional near-infrared spectroscopy signals. *Journal of Biomedical Optics, 16*(10), 106005.

Witteman, J., Van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: a quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia, 50*, 2752–2763.

Zhang, D., Chen, Y., Hou, X., & Wu, Y. J. (2019). Near-infrared spectroscopy reveals neural perception of vocal emotions in human neonates. *Human Brain Mapping, 40*(8), 2434–2448.

Zhang, D., Zhou, Y., Hou, X., Cui, Y., & Zhou, C. (2017). Discrimination of emotional prosodies in human neonates: A pilot fNIRS study. *Neuroscience Letters, 658*, 62–66.

Zhang, D., Zhou, Y., & Yuan, J. (2018). Speech prosodies of different emotional categories activate different brain regions in adult cortex: an fNIRS study. *Scientific Reports, 8*(1), 218.

Notes

[2] STC comprises the superior temporal gyrus (STG), middle temporal gyrus (MTG), and superior temporal sulcus (STS).

Note: Figure translations are in progress. See original paper for figures.

ChinaRxiv 中X

Source: ChinaXiv —Machine translation. Verify with original.