Neurosci. Bull. October, 2018, 34(5):801–815
https://doi.org/10.1007/s12264-018-0231-9

www.neurosci.cn
www.springer.com/12264

ORIGINAL ARTICLE

Facial Expression Enhances Emotion Perception Compared to Vocal Prosody: Behavioral and fMRI Studies

Heming Zhang[1] · **Xuhai Chen**[2,3] · **Shengdong Chen**[1] · **Yansong Li**[4] ·
Changming Chen[5] · **Quanshan Long**[1] · **Jiajin Yuan**[1]

Received: 16 December 2017 / Accepted: 13 March 2018 / Published online: 9 May 2018

Abstract Facial and vocal expressions are essential modalities mediating the perception of emotion and social communication. Nonetheless, currently little is known about how emotion perception and its neural substrates differ across facial expression and vocal prosody. To clarify this issue, functional MRI scans were acquired in Study 1, in which participants were asked to discriminate the valence of emotional expression (angry, happy or neutral) from facial, vocal, or bimodal stimuli. In Study 2, we used an affective priming task (unimodal materials as primers and bimodal materials as target) and participants were asked to rate the intensity, valence, and arousal of the targets. Study 1 showed higher accuracy and shorter response latencies in the facial than in the vocal modality for a happy expression. Whole-brain analysis showed enhanced activation during facial compared to vocal emotions in the inferior temporal-occipital regions. Region of interest analysis showed a higher percentage signal change for facial than for vocal anger in the superior temporal sulcus. Study 2 showed that facial relative to vocal priming of anger had a greater influence on perceived emotion for bimodal targets, irrespective of the target valence. These findings suggest that facial expression is associated with enhanced emotion perception compared to equivalent vocal prosodies.

Keywords fMRI · Emotion perception · Facial expression · Vocal prosody · Modality

Introduction

Emotion is the core of social relations [1, 2]. Successful interpersonal communication in the social environment requires an exact understanding of people’s feelings, thoughts, and intentions. The key to understanding the minds of others is the ability to surmise complex mental states from slight sensory cues [3]. In daily life, we usually perceive emotions through two aspects: verbal and non-verbal emotional stimuli [4, 5]. Among these, nonverbal emotional communication is often carried out through different modalities such as vocal prosodies and facial expressions. Although we often perceive emotions from multiple modalities in social interactions, we can also rely on a single modality (visual or auditory) to perceive the emotions of others [6, 7]. In fact, it is easy to perceive happiness when we look at a picture of a face with a big smile.

Previous studies investigated emotion perception mainly by focusing on the difference between emotion categories within one modality, either facial expressions or vocal prosodies. Early studies using electrical stimulation, lesion, and microelectrode recording analysis found several brain regions involved in facial affect processing including the

✉ Jiajin Yuan
yuanjiajin168@126.com; yuanjiaj@swu.edu.cn

[1] Key Laboratory of Cognition and Personality of the Ministry of Education, School of Psychology, Southwest University, Chongqing 400715, China

[2] Key Laboratory of Behavior and Cognitive Psychology in Shaanxi Province, School of Psychology, Shaanxi Normal University, Xi’an 710062, China

[3] Key Laboratory of Modern Teaching Technology of the Ministry of Education, Shaanxi Normal University, Xi’an 710062, China

[4] Department of Psychology, School of Social and Behavioral Sciences, Nanjing University, Nanjing 210023, China

[5] School of Educational Sciences, Xinyang Normal University, Xinyang 464000, China

amygdala [8–11], right temporal lobe [12, 13], right or left basal ganglia and anterior temporal lobe [14], and the right mesial occipital and right inferior parietal regions [15]. In contrast, studies using positron emission tomography, magnetoencephalography, and functional magnetic resonance imaging (fMRI) found evidence supporting the existence of prosody-specific right hemisphere processing [16–19], especially activation in the lateral temporal lobes [19], right prefrontal cortex [20], and right frontal and temporal regions [21] during the processing of emotional prosody. For example, studies using facial expressions found positive correlations between perceived anger ratings and the intensity of angry facial expressions, whose intensity increase was linked with enhanced activity in the orbitofrontal and anterior cingulate cortex [4]. Gur *et al.* (1994) found that a happy relative to a sad discrimination task activated the left frontal region [22]. Some studies found more activation during happy relative to neutral facial expression in the left anterior amygdala, bilateral fusiform [23], left anterior cingulate gyrus, bilateral posterior cingulate gyri, medial frontal cortex, right supramarginal gyrus [24], and medial frontal/cingulate sulcus region [25], while more activation was found during angry than neutral facial expressions in the posterior part of the right cingulate gyrus, left medial temporal gyrus [26], bilateral fusiform gyri, and left inferior frontal gyrus [25]. In addition, increasing the intensity of a happy facial expression enhanced activity in the bilateral fusiform gyri and right putamen [7]. These studies investigated the different valences of emotions perceived only through facial expressions.

On the other hand, using the same valences of emotion through vocal prosodies only, studies have reported more activation during happy than angry prosodies in the right anterior middle temporal gyrus, right inferior frontal gyrus, and bilateral posterior middle temporal gyrus [27]. Moreover, it was reported that angry prosodies induced more activation than neutral prosodies in the middle superior temporal sulcus [28] while unpleasant prosodies induced more activation than pleasant prosodies in the amygdala, hippocampus, parahippocampal gyrus, and temporal poles [29]. Also, Ethofer *et al.* [5] reported faster reaction times for angry than for neutral prosodies, which was coupled with enhanced activation during angry *versus* neutral prosodies in the voice-sensitive temporal cortices, amygdala, insula, and mediodorsal thalami during an emotion classification task [5].

The above evidence confirms first, that different modalities can express the same emotion, and second, that people can rely on a single modality to perceive the emotions of others. However, it has yet to be determined how the perception of a given emotional expression and its neural substrate varies from the visual to the auditory modality. Hulka *et al.* used the Comprehensive Affect Testing System to measure emotion perception across the channels of facial affect and prosody in cocaine users [30]. The results showed that cocaine users had significantly lower scores than healthy controls in the quotient scales of “prosody recognition” whereas facial affect perception was comparable in both samples [30]. Chen *et al.* have shown that P3, alpha, and beta activity are significantly smaller for the neutral condition than for facial-anger and bimodal-anger, but not for vocal-anger over all brain regions [31]. These results suggest that the emotion perception effect of the visual modality is stronger than that of the auditory modality [31]. However, this study did not directly compare the emotion perception across modalities. To address this issue, we focused on the modality effect in emotion perception between facial expression and vocal prosodic materials. Based on these findings, we assumed that the presentation of facial expression is associated with enhanced emotion perception in behavioral and neuroimaging measures compared to the equivalent vocal prosodies.

In the current studies, we controlled the intensity, valence, and arousal of the stimuli across different modalities, in order to reliably compare the emotion perception between facial and vocal modalities. In addition, visual and auditory materials differ in many ways. To increase the comparability of facial and vocal expression and to attribute the differences between facial and vocal stimuli solely to cross-modal emotion perception, we imbedded facial and vocal expression into the same neutral, non-emotional audiovisual stimulus context. Another advantage of this method is to increase the ecological validity of laboratory emotion perception. Emotional expression in real-life settings, in most cases, unfolds dynamically, such as anger or happiness evolving from a neutral, non-emotional state. Both behavioral [32–34] and neuroimaging evidence [35–38] suggests that dynamic facial expression conveys emotion that more resembles real-life facial communication than static expression (e.g. a photograph). Based on these considerations, the emotional expressions used in our studies were manipulated to be dynamic rather than static materials. In Study 1 we collected fMRI scans when participants were required to label the category of emotional expression. However, the affect recognition task in Study 1 did not allow direct assessment of the perceived emotion intensity, and neuroimaging results from occipital or temporal regions may be related to modality-specific function. In order to remove modality-specific effects and verify the conclusion of Study 1, in Study 2 we used an affective priming task, which directly assessed the experiential emotion intensity for bimodal target stimuli following the unimodal primes. Together, we performed these two studies to investigate emotion perception differences between visual and auditory modalities.

Study 1

Materials and Methods

Participants

Twenty-five right-handed university students (13 women, aged 19–24 years, mean 20.85) were recruited to participate in the experiment. All participants reported normal auditory and normal or corrected-to-normal visual acuity, and were free of neurological or psychiatric problems. They were emotionally stable. The experimental procedure adhered to the Declaration of Helsinki and was approved by the Local Review Board for Human Participant Research. All participants gave informed consent prior to the study and were reimbursed ¥60 for their time.

Stimuli

The facial stimuli were neutral, angry, and happy facial expressions of a male actor from the set created by the Chinese Facial Affective Picture System [39]. The pictures were cropped such that non-facial body parts were not visible.

The vocal stimuli were 15 sentences in Standard Chinese of neutral content (each sentence lasting for 3.65 s) produced by one trained male actor in neutral, happy, and angry prosodies which have been validated in previous studies [40, 41]. Following the previous studies [31, 42], sentences with vocal emotion were obtained by cross-splicing the first part of a neutral prosody and the second part of an emotional prosody (Fig. 1). Then, each sentence was paired with a facial expression. Consistent with the structure of the sentences, the facial expressions varied at the splicing point from neutral to angry or happy while the identity was kept unchanged (Fig. 1). Four types of facial-vocal stimuli, that is, no emotion, vocal-emotion, facial-emotion, and bimodal-emotion, were established. Trials were presented randomly within each block, such that the emotion category itself could not be predicted before stimulus onset. And because we were particularly interested in facial-vocal comparison in emotion effects, we controlled the valence (ranging from unpleasant to pleasant), arousal (from calm to excited), and intensity (the extent of emotional display by an actor) between facial and vocal emotional conditions.

Design and Procedure

Participants performed two fMRI runs, each starting and ending with a 10-s fixation baseline. Within each run, 90 trials were presented in 6 blocks of 15 trials. These blocks were separated by a 5-s fixation period. The 3 blocks differed in the type of stimulus presented (facial, vocal, or bimodal emotion), and each included one neutral and two emotion (anger, happiness) conditions. It is noteworthy that all the stimuli were presented through visual and auditory modalities. For example, in the facial-emotion block, the facial expression varied at the splicing point from neutral to angry or happy while the neutral prosody remained unchanged. Correspondingly, in the vocal-emotion block, the prosody varied at the splicing point from neutral to angry or happy while the neutral facial expression remained unchanged. And in the bimodal-emotion block, both the facial expression and the prosody varied at the splicing point from neutral to angry or happy. Trials were presented in a random order within each block. Each run lasted for 14 min and 5 s.

Each trial consisted of a 2-s fixation cross, followed by a 3.65-s stimulus, a 1-s blank screen, and a 2.35-s response window. During the presentation of each stimulus, participants were asked to judge the valence of the emotional

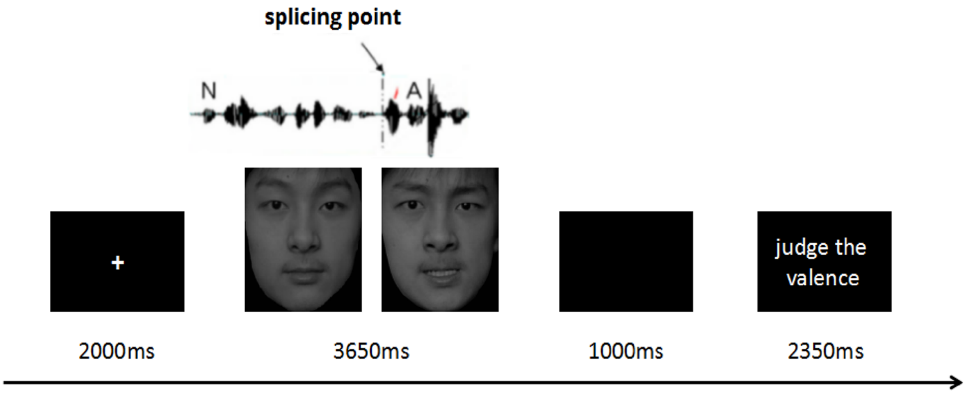

Fig. 1 An example of a bimodal-emotion trial under the anger condition. Trials consisted of fixation, stimulus presentation, blank, and judgment. The lower stimuli represent neutral and angry facial expressions, while the upper stimuli represent neutral and angry vocal prosodies. Facial and vocal emotional expressions were equalized for onset time and duration. Four types of facial-vocal stimuli (no-emotion, vocal-emotion, facial-emotion, and bimodal-emotion) were established.

expression on a 3-point scale (1 “positive”, 2 “negative”, 3 “neutral”, counterbalanced across participants) on a keypad held in the right hand. Before the experiment, we asked participants to focus on the prosody rather than the semantics of the speaker. This task was chosen to encourage participants to actively perceive emotions through diverse modalities and to avoid potential distraction of attention. Participants were instructed to classify the stimuli as accurately and quickly as possible, but withhold their motor response until presentation of the response window.

Data Acquisition

Behavioral Analysis

For statistical convenience, we calculated accuracy (ACC) as a decimal. Response times (RTs) beyond three standard deviations and with wrong responses were excluded. The descriptive results of ACC and RTs are depicted in Fig. 2. Behavioral data were analyzed with SPSS 22 (SPSS 22.0 for Windows) (IBM, Armonk, NY). Correlation analyses and repeated measures analysis of variance (ANOVA) were performed for ACC and RTs.

fMRI Acquisition and Analysis

Brain images were acquired with a Siemens 3 Tesla scanner (Magnetom Trio, Siemens, Erlangen, Germany). Whole-brain blood oxygenation-level dependent (BOLD) functional images were collected with a gradient echoplanar imaging sequence [repetition time (TR) = 2000 ms; echo time (TE) = 25 ms; flip angle = 75°; matrix size = 64 × 64; field of view (FoV) = 220 × 220 mm^2; voxel size = 3.4 × 3.4 × 3 mm^3]. Each functional run contained 428 brain volumes, and each volume comprised 32 axial slices. T1-weighted anatomical images were recorded with a total of 176 slices at a thickness of 1 mm (TR = 1900 ms; TE = 2.52 ms; flip angle = 9°, matrix = 64 × 64, FoV = 256 × 256 mm^2, voxel size = 1 × 1 × 1 mm^3). Stimulus presentation and behavioral data were acquired using E-prime software (Psychology Software Tools, Inc, Sharpsburg, GA).

The fMRI data were preprocessed using DPABI (The R-fMRI Network (RFMRI.ORG), Institute of Psychology, Chinese Academy of Sciences) [43] and comprised slice-timing, spatial realignment to correct for head movement during the scanning, and nonlinear warping into Montreal Neurologic Institute (MNI) space using unified segmentation on T1 images. Normalized functional images were re-sampled to 3 × 3 × 3 mm^3 voxels and spatially smoothed with a Gaussian kernel of 6-mm full-width at half maximum. Head movement estimates from the realignment step were included as regressors in all analyses to help diminish the impact of any movement-related effects on the results.

Statistical analysis of the preprocessed functional data was performed in statistical parametric mapping (SPM8; www.fil.ion.ucl.ac.uk/spm) and programs custom-written in MatLab (MathWorks, Natick, MA). For the purpose of the present analysis, we focused on neural responses while the participants were watching and listening to the stimuli. For each participant, a voxel-wise whole-brain analysis was implemented using the general linear model. Nine periods of interest (facial anger, vocal anger, bimodal anger, facial happiness, vocal happiness, facial neutral, vocal neutral, bimodal neutral and bimodal happiness) were included in the model to compute linear contrast maps. Six realignment parameters were further included as regressors of no interest to account for head motion effects. The group-level statistical parametric maps were produced according to repeated measures ANOVA to enhance the generalizability of the results. Our preliminary analysis showed that the effects of primary interest (i.e., Emotion (3) × Modality (3) interaction) were not affected by the accuracy of discrimination, as verified by the lack of significant Emotion, Modality and Accuracy interaction in the whole-brain analysis ($P_{\mathrm{FWE}} > 0.99$). Thus, we included all the trials for fMRI analyses to obtain a better quality of signals. Simple effects paired *t*-tests were conducted to investigate brain activations that had a linear relationship as a function of the interaction of modality and emotion. The whole-brain regression analyses used a significance threshold of $P_{\mathrm{FWE}} < 0.05$ with a 10-voxel extent threshold [peak family-wise error (FWE) correction].

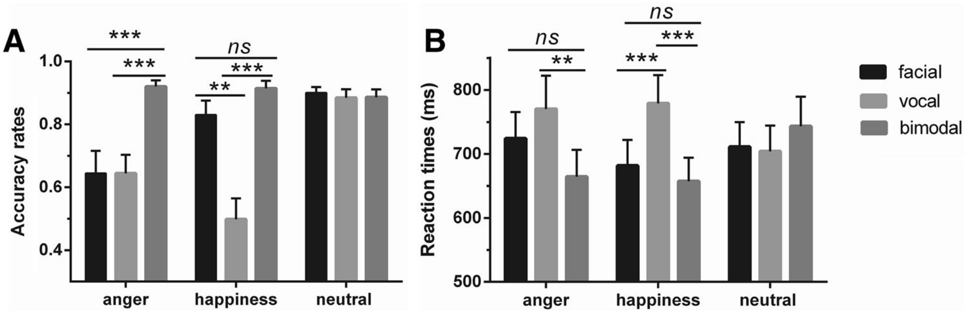

Fig. 2 Statistics of behavioral performance. **A** Accuracy rates; **B** reaction times. Error bars indicate standard error (**$P \le 0.01$, ***$P \le 0.001$; *ns*, no significant difference).

The major goal of this fMRI study was the detailed characterization of the response profile across experimental conditions for the modality effect in emotion perception. In this context, ROI analysis was next conducted. Peelen *et al.* have suggested that the medial prefrontal cortex (MPFC) and superior temporal sulcus (STS) represent perceived emotions at an abstract, modality-independent level [3], so we set them as spherical ROIs. ROI analyses of the STS and MPFC were defined anatomically based on MNI coordinates (MPFC: $x = 11$, $y = 49$, $z = 21$; STS: $x = -47$, $y = -64$, $z = 5$) from a prior analysis [3] using an 8-mm radius. The percentage signal change (PSC) for each ROI was then extracted using MarsBaR [44]. The modality effects in emotion perception were operationally defined by the differences in PSC values between facial and vocal emotion displays.

Results

Behavioral Performance

Participants rated the valence of the perceived emotions on a 3-point scale (1 “positive”, 2 “negative”, 3 “neutral”, counterbalanced across participants). A 3 (Emotion) × 3 (Modality) repeated measures ANOVA of the ACC data revealed a significant main effect of Modality ($F(1, 32) = 14.139$, $P < 0.01$, $\eta^2 = 0.371$) and Emotion ($F(2, 48) = 15.895$, $P < 0.01$, $\eta^2 = 0.398$) and a significant interaction between Emotion and Modality ($F(2, 59) = 12.143$, $P < 0.01$, $\eta^2 = 0.336$). Bonferroni *post hoc* tests showed lower vocal ACC than facial ($P < 0.01$) and bimodal ($P < 0.001$) ACC during the perception of a happy expression, while the differences between bimodal and facial modalities were not significant ($P > 0.05$). For the angry expression, the bimodal ACC was higher than that for the vocal ($P < 0.001$) and facial modalities ($P < 0.001$).

The repeated measures ANOVA for RTs yielded a significant main effect of Modality ($F(2, 48) = 5.601$, $P < 0.01$, $\eta^2 = 0.189$) and a significant interaction between Emotion and Modality ($F(4, 96) = 8.551$, $P < 0.01$, $\eta^2 = 0.263$). Irrespective of the emotion category, the RTs were shorter for the bimodal than for the vocal modality ($P < 0.01$), while the RTs were similar for the bimodal and facial modalities ($P > 0.05$). In addition, the RTs were shorter for the facial than for the vocal modality ($P < 0.001$) specifically during a happy expression (Fig. 2).

Emotional Ratings

In a pilot study, 30 participants (mean age = 21.37 ± 1.52 years) who did not participate in the experiment were recruited to rate the valence, arousal, and intensity of all the stimulus materials using a 9-point scale. After data collection, we calculated one-way ANOVAs of the valence, arousal, and intensity rating data. Bonferroni *post hoc* tests were performed after detecting significant main effects of emotion category and modality (Fig. 3). The results revealed that all the materials of unimodal emotion

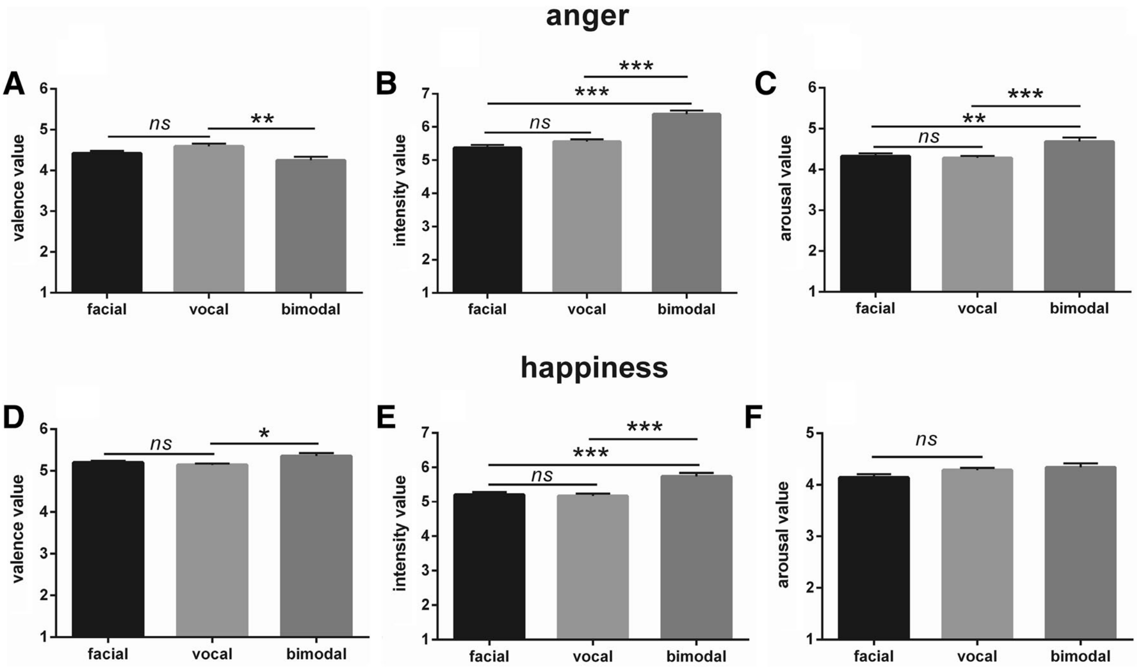

Fig. 3 Rating data of valence, intensity, and arousal. **A–C**: valence, intensity, and arousal values for anger expression. **D–F**: valence, intensity, and arousal values for happiness expression. *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$; *ns*, no significant difference.

and those of bimodal emotion were rated significantly different from the neutral materials in intensity, valence, and arousal ($P < 0.02$). Moreover, the materials of facial and vocal emotion were rated similar in intensity, valence, and arousal ($P > 0.1$). The materials of bimodal emotion were rated higher than those of either unimodal condition in intensity and arousal for an angry expression ($P < 0.01$)(Fig. 3).

Results of Whole-Brain Analysis

Main Effect of Modality and Emotion

The repeated measures ANOVAs showed that the main effect of modality ($F(2,192) = 13.333$, $P_{\mathrm{FWE}} < 0.05$, voxel > 10) activated a neural network including the bilateral inferior temporal gyrus, bilateral middle occipital gyrus, right fusiform gyrus, and right inferior occipital gyrus, as well as the right middle temporal gyrus (Table 1). The main effect of emotion ($F(2,192) = 13.333$, $P_{\mathrm{FWE}} < 0.05$, voxel > 10) activated an emotion processing-related neural network that included the bilateral inferior frontal gyrus, bilateral superior temporal gyrus, bilateral middle temporal gyrus, bilateral lingual gyrus, and right middle frontal gyrus (Table 1).

Interaction of Modality and Emotion

The interaction effect of modality and emotion ($F(4,192) = 8.319$, $P_{\mathrm{FWE}} < 0.05$, voxel > 10) showed activation in an emotion processing network that included the bilateral fusiform gyrus, bilateral inferior occipital gyrus,

Table 1 Main effects of emotion and modality generated by 3 (Emotion) × 3 (Modality) repeated measures ANOVA.

Brain regions	MNI coordinates			Cluster size	*F*-value
	x	y	z		
Main effect of modality					
R middle occipital gyrus	39	−81	9	60	$F(2,192) = 20.18$
	48	−75	−3		$F(2,192) = 15.10$
L middle occipital gyrus	−48	−72	−3	26	$F(2,192) = 17.23$
	−38	−81	6		$F(2,192) = 13.61$
R inferior temporal gyrus	51	−72	−6	58	$F(2,192) = 18.88$
	42	−57	−15		$F(2,192) = 17.36$
L inferior occipital gyrus	−45	−72	−9	37	$F(2,192) = 17.68$
R middle temporal gyrus	51	−66	−3	42	$F(2,192) = 18.93$
R fusiform gyrus	36	−54	−21	70	$F(2,192) = 18.47$
	33	−29	−24		$F(2,192) = 16.23$
Main effect of emotion					
R superior temporal gyrus	51	−33	3	264	$F(2,192) = 33.37$
	54	−15	−6		$F(2,192) = 30.99$
	60	−6	−9		$F(2,192) = 27.76$
R middle temporal gyrus	51	−30	0	85	$F(2,192) = 29.37$
	51	−15	−12		$F(2,192) = 20.92$
	54	−24	−6		$F(2,192) = 20.67$
L middle temporal gyrus	−57	−21	−3	74	$F(2,192) = 24.88$
L inferior frontal gyrus, orbital part	−42	27	−9	27	$F(2,192) = 16.64$
	−30	21	−18		$F(2,192) = 16.32$
R inferior frontal gyrus, triangular part	48	30	0	69	$F(2,192) = 21.18$
L inferior frontal gyrus, triangular part	−51	21	6	26	$F(2,192) = 16.93$
	−48	27	0		$F(2,192) = 14.82$
R middle frontal gyrus	27	39	33	12	$F(2,192) = 15.27$
L calcarine fissure	−12	−81	3	42	$F(2,192) = 25.21$
R lingual gyrus	12	−63	−9	39	$F(2,192) = 26.98$
L lingual gyrus	−9	−81	0	67	$F(2,192) = 25.85$

MNI coordinates and F-statistics are listed for the voxel within each cluster with the maximum BOLD response (all clusters $P_{\mathrm{FWE}} < 0.05$, peak FWE correction). MNI, Montreal Neurological Institute; BOLD, blood oxygenation level-dependent; L, left; R, right.

bilateral inferior temporal gyrus, bilateral middle temporal gyrus, bilateral middle occipital gyrus, bilateral superior temporal gyrus, and right temporal pole (Table 2, Fig. 4). To differentiate between modalities, the simple effects were analyzed separately for angry, happy, and neutral conditions ($T(24) = 5.696$, $P_{\mathrm{FWE}} < 0.05$, voxel > 10). For the anger condition, the facial emotion elicited greater activation than did the vocal emotion in the right inferior temporal gyrus and left middle occipital gyrus, while bimodal-emotion elicited greater activation than vocal-emotion in the right fusiform gyrus, left middle occipital gyrus, and left inferior occipital gyrus. For the happiness condition, the activations in the right inferior temporal gyrus and right fusiform gyrus were greater in response to facial-emotion than to vocal-emotion, while the activations in the right middle occipital gyrus and right fusiform gyrus were greater in response to bimodal-emotion than to vocal-emotion. In addition, activations in the right fusiform gyrus and right middle occipital gyrus were greater in response to bimodal-emotion than to vocal-emotion.

However, the contrast of bimodal relative to facial emotion did not show any significantly activated regions, irrespective of emotion category; and the vocal relative to facial contrast exhibited no activated areas either. Also, the cross-modal contrasts did not show significantly activated areas for the neutral condition.

Region of Interest Analysis

A 3 (Emotion) × 3 (Modality) repeated measures ANOVA of the PSCs in the STS revealed a significant main effect of modality ($F(2, 38) = 4.814$, $P < 0.02$, $\eta^2 = 0.167$) and a significant interaction between Emotion and Modality ($F(4, 96) = 4.728$, $P < 0.01$, $\eta^2 = 0.165$). The *post hoc* comparisons with the Bonferroni method showed higher PSCs for facial emotion than for vocal emotion ($P < 0.01$) during the anger condition, whereas facial and bimodal anger expression elicited similar STS activity ($P > 0.05$). The analysis of PSCs during the happy condition did not show significant differences across the facial, vocal, and bimodal

Table 2 The 3 (Emotion) × 3 (Modality) interaction effects generated by repeated measures ANOVA.

Brain regions	MNI coordinates			Cluster size	*F*-Value
Interaction	x	y	z		
R superior temporal gyrus	60	−6	−6	172	$F(4,192) = 16.60$
L superior temporal gyrus	−63	−27	3	53	$F(4,192) = 13.01$
	−57	−12	−3		$F(4,192) = 12.96$
R inferior temporal gyrus	42	−60	−12	136	$F(4,192) = 20.10$
	48	−63	−6		$F(4,192) = 18.92$
	45	−54	−24		$F(4,192) = 10.57$
R middle temporal gyrus	51	−66	0	123	$F(4,192) = 20.93$
	45	−75	9		$F(4,192) = 10.14$
L middle temporal gyrus	−48	−69	0	10	$F(4,192) = 17.53$
	−60	−15	−3	108	$F(4,192) = 14.49$
	−63	−30	3		$F(4,192) = 13.10$
R temporal pole, superior temporal gyrus	57	3	−12	25	$F(4,192) = 13.11$
L inferior occipital gyrus	−48	−78	−6	91	$F(4,192) = 24.66$
R inferior occipital gyrus	42	−63	−12	88	$F(4,192) = 19.30$
	45	−75	−3		$F(4,192) = 17.49$
L middle occipital gyrus	−48	−75	−3	156	$F(4,192) = 25.80$
	−42	−87	−9		$F(4,192) = 12.51$
R middle occipital gyrus	42	−78	0	76	$F(4,192) = 17.69$
	30	−90	9		$F(4,192) = 9.03$
R fusiform gyrus	42	−60	−15	130	$F(4,192) = 18.45$
	39	−51	−18		$F(4,192) = 18.13$
	36	−69	−15		$F(4,192) = 13.04$
L fusiform gyrus	−42	−81	−15	84	$F(4,192) = 13.66$
	−39	−54	−21		$F(4,192) = 12.95$

MNI coordinates and F-statistics are listed for the voxel within each cluster with the maximum BOLD response (all clusters $P_{\mathrm{FWE}} < 0.05$, peak FWE correction). MNI, Montreal Neurological Institute; BOLD, blood oxygenation level-dependent; L, left; R, right.

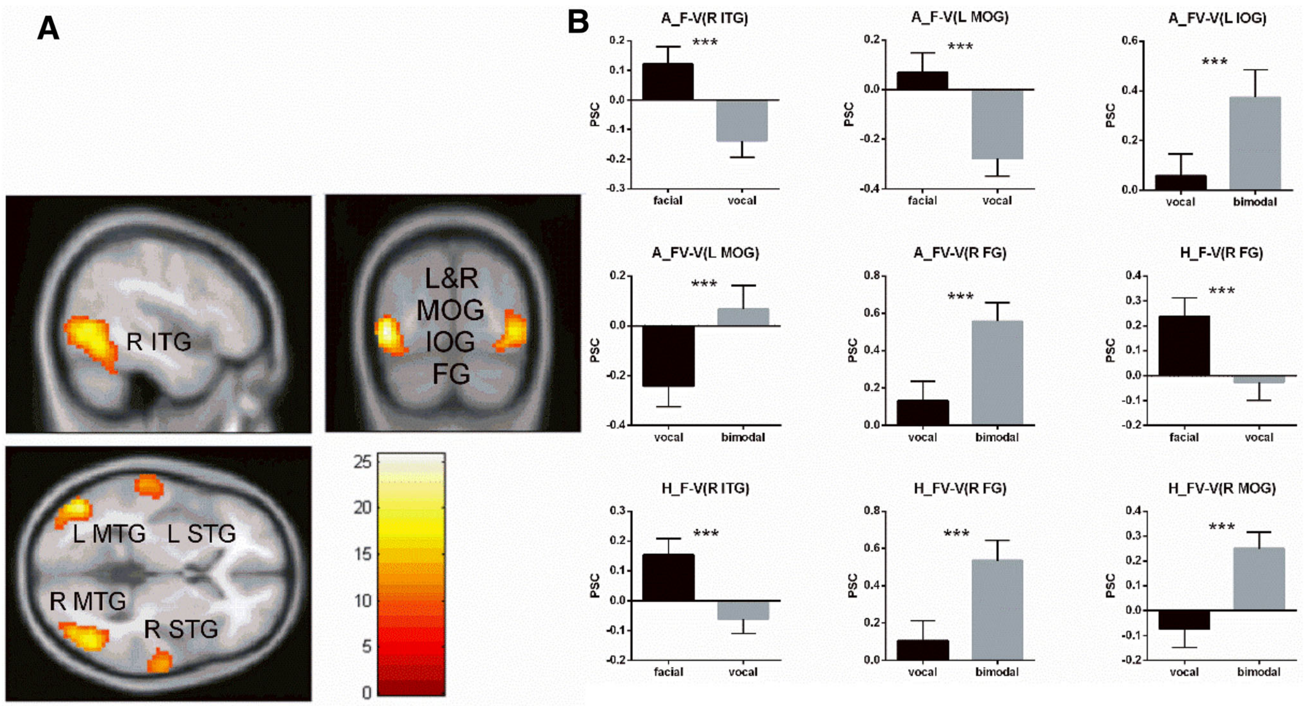

Fig. 4 **A** Areas activated by the Emotion by Modality interaction in whole-brain analysis (i.e., L and R FG, L and R IOG, R ITG, L and R MTG, Land R MOG, L and R STG, and R temporal pole; all clusters $P_{\mathrm{FWE}} < 0.05$). **B** Histograms of significant simple effects ($P < 0.001$). The contrast of facial *vs* vocal anger elicited activations in the L MOG and the R ITG, and the contrast of facial *vs* vocal happiness elicited activations in the R FG and R ITG. Error bars indicate standard error. A_F-V (H_F-V), facial *vs* vocal contrast during anger (happiness) expression; A_FV-V (H_FV-V), bimodal *vs* vocal contrast during anger (happiness) expression. ITG, inferior temporal gyrus; FG, fusiform gyrus; MOG, middle occipital gyrus; IOG, inferior occipital gyrus; MTG, middle temporal gyrus; STG, superior temporal gyrus; L, left; R, right.

conditions ($P > 0.05$). By contrast, the results of PSC analysis in the MPFC did not show significant differences between facial and vocal modalities, irrespective of emotion category ($P > 0.05$, Fig. 5).

Therefore, this fMRI study showed that facial relative to vocal prosodies recruited more intense neural processing in the occipital-temporal network (e.g. the right inferior temporal gyrus, left middle occipital gyrus, and STS). However, although STS is associated with the modality-independent representation of emotional stimuli [3], occipital and inferior temporal network areas are well known for their roles in visual selective attention [45–48]. That is, the enhanced perceptual processing of facial *versus* prosodic stimuli in these regions may result from their specific roles in visual processing. In order to test whether our conclusion of enhanced emotion perception during facial relative to prosodic stimuli is reliable and independent of modality-specific effects, we conducted an affective-priming paradigm in Study 2 wherein bimodal emotional expression served as the target and unimodal expression as the priming stimulus. If the emotion perception of bimodal expression is enhanced following facial relative to prosodic priming, it can be reliably concluded that facial expression is linked with more enhanced emotion perception than vocal prosody.

Study 2

Materials and Methods

Participants

Thirty-three right-handed university students (mean age 19.67 ± 1.47) were recruited. All participants reported normal auditory and normal or corrected-to-normal visual/auditory acuity, and were free of neurological or psychiatric problems. They self-reported that they were emotionally stable. The experimental procedure adhered to the Declaration of Helsinki and was approved by the Local Review Board for Human Participant Research. All participants gave informed consent prior to the study and were reimbursed ¥30 for their time.

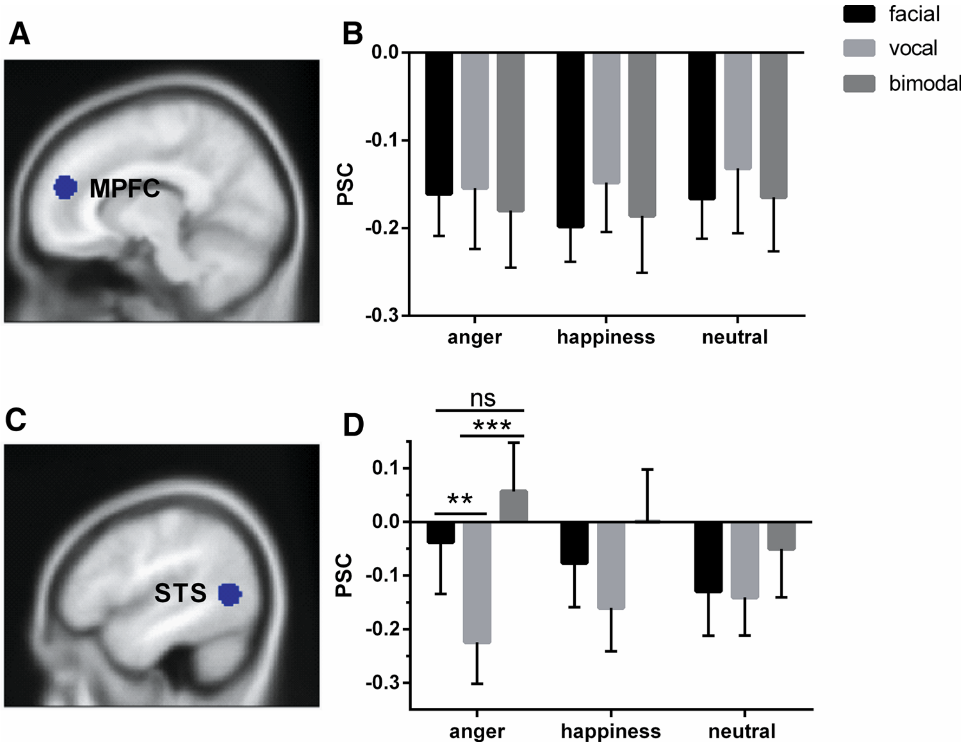

Fig. 5 ROI analysis in the MPFC and STS. **A** MPFC; center of sphere (x, y, z): 11, 49, 21, r = 8 mm; **B** PSCs were consistent across different emotional and modality conditions. **C** STS; center of sphere (x, y, z): −47, −64, 5, r = 8 mm; **D** PSCs were lower for the vocal than for the bimodal and the facial modality during anger expression, while there was no significant difference between facial and bimodal anger expression (**$P \le 0.01$, ***$P \le 0.001$; *ns*, no significant difference).

Stimuli

As in Study 1.

Behavioral Procedure

Participants performed 2 blocks. Each block included 135 trials, each including a unimodal primer and a bimodal target stimulus. The 2 blocks differed in the type of priming stimulus (facial or vocal emotion), and each block consisted of three conditions (neutral, anger, or happiness). Similarly, there were three conditions—neutral, angry, or happy expression—for the bimodal target stimulus. Trials were presented in a random order within each block.

Each trial started with a 300-ms fixation cross, followed by a 3.65-s priming stimulus. After a blank screen for 1 s, a target was presented for 4 s. Then, a blank screen was presented for 1 s and was followed by a 3-s response window. After the target presentation, participants were asked to assess the valence, arousal, and intensity of the bimodal target expression on a 9-point scale on a keypad held in the right hand.

Data Analysis

The descriptive results of rating data of valence, intensity, and arousal are shown in Fig. 6. The data analysis method in Study 2 was similar to the behavioral data analysis in Study 1.

Results

A 3 (priming emotion) × 2 (priming modality) × 3 (target emotion) repeated measures ANOVA of valence, arousal, and intensity rating data was carried out. Bonferroni *post hoc* tests were performed after detecting significant interaction effects of priming emotion, priming modality, and target emotion (valence: $F(2, 66) = 7.279$, $P < 0.01$, $\eta^2 = 0.185$; intensity: $F(2, 71) = 16.047$, $P < 0.001$, $\eta^2 = 0.334$; arousal: $F(3, 88) = 7.581$, $P < 0.001$, $\eta^2 = 0.192$; Fig. 6). The results showed that for an angry target preceded by an angry primer, the intensity and arousal ratings were significantly lower but the valence rating was higher during prosodic than during facial priming (*Ps* < 0.01). For a happy target with angry primers, the intensity and valence ratings were significantly higher (i.e., more pleasant) but the arousal rating was lower during prosodic than during facial priming (*Ps* < 0.01), suggesting that negative facial primers interfere with an individual's positive rating for the happiness target more than the corresponding prosodic primer. For neutral targets, there was no significant difference in these values between the two priming conditions (Fig. 6).

Thus, based on all the valence, arousal, and intensity ratings, the results of Study 2 consistently showed that emotional perception of bimodal anger was enhanced while that of bimodal happiness was hampered following facial relative to prosodic anger priming. This further verifies that it is a reliable phenomenon that facial expression results in

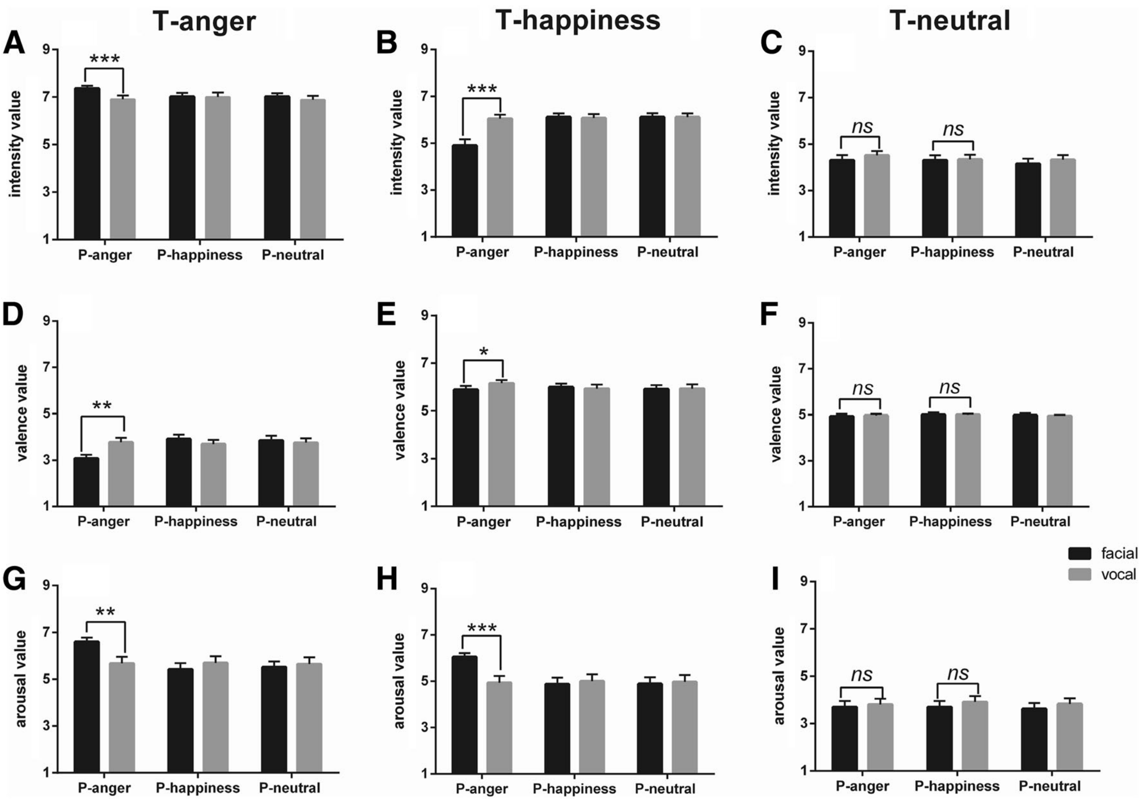

Fig. 6 Intensity (**A–C**), valence (**D–F**), and arousal (**G–I**) ratings. T, target; P, primer. Error bars indicate standard error (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$; *ns*, no significant difference).

more intense emotional perception than the equivalent vocal prosody.

Discussion

Using a valence judgment task and three rating tasks, we aimed to test whether the presentation of facial expressions is associated with enhanced emotional perception compared to equivalent vocal prosodies. As predicted, our behavioral, whole-brain and ROI analyses consistently showed that facial expression recruited faster and more accurate perception of emotional display, and recruited more intense neural processing in the occipital-temporal network (e.g. the right inferior temporal gyrus, left middle occipital gyrus, and STS) than vocal prosodies. In order to control modality-specific effects, we conducted Study 2 using an affective priming and emotion rating task. The results showed a greater influence of negative facial than vocal priming on emotion ratings for bimodal targets, irrespective of valence. This is consistent with the neuroimaging results of Study 1 in showing enhanced emotional perception during facial relative to prosodic stimulation.

First, our judgment results showed higher ACC and shorter RTs in facial than in vocal emotion for a happy expression, but similar ACC and RTs for facial and vocal emotions with angry and neutral expressions. By contrast, whole-brain analysis of the neuroimaging data showed higher activation for facial than vocal emotion in the inferior temporal-occipital network during an angry expression. Thus, there appears to be a dissociation between behavioral and neuroimaging data concerning the emotion category exhibiting enhanced perception of facial *vs* vocal emotion. Specifically, the phenomenon of enhanced perception of facial *vs* vocal emotion was more prominent for happiness in the behavioral data, but more prominent for anger in the neuroimaging data. Previous studies have reported a happy face advantage during the recognition of happy relative to negative facial expressions, due to the unique and distinctive features of a happy facial expression (e.g. a smiling mouth; [49, 50]). For example, earlier behavioral studies have suggested that facial expressions of happiness are categorized faster than those of negative emotions such as anger [51] and sadness [52]. Recently, using a visual search task, Farran and colleagues reported that the identification of angry, sad, and fearful facial expressions was impaired in high-functioning autistic patients while their recognition of happy facial expressions was intact compared to a healthy population [53].

However, as a result of emotional negativity bias, the emotion processing in the brain favors negative over positive information, though this sensitivity is sometimes not apparent in behavioral measures [54–56]. For example, Sprengelmeyer *et al.* reported that the intensity of angry facial expressions does not have any significant effect on RTs or error rates, while the N170 and the 200–600 ms event-related potential (ERP) amplitudes increase with the intensity of an angry expression in the inferior temporal-occipital area [57]. In addition, Leppänen *et al.* reported faster RTs for happy than fearful facial expressions while brain potentials were sensitive to the intensity of fearful but not happy expressions at occipital-temporal sites [54]. It is worth noting that the neural sensitivity to negative facial expressions in these ERP studies was recorded in the inferior temporal-occipital region, consistent with the current finding that the inferior temporal gyrus and middle occipital gyrus were more sensitive in detecting the enhanced perception of facial *vs* vocal anger expression. The role of the inferior temporal-occipital network in mediating this modality effect is further supported by our finding of greater activation of the right inferior temporal gyrus and fusiform gyrus during the facial *vs* vocal perception of happiness. Consistent with the theory of a happy face advantage [49, 58], it has been reported that detecting a single distinctive feature of a happy face is sufficient to activate face-sensitive areas like the fusiform gyrus [59]. However, this distinctive feature is absent when happiness is displayed by vocal prosody, thus leading to the enhanced activation of face-sensitive areas during happy face perception.

In the whole-brain analysis, the facial *versus* vocal contrast showed reliable activation in inferior temporal-occipital regions. There are two possible explanations for these results. First, facial expression is a major signal on which humans rely to decode others' emotional states. Exact face-based emotion recognition and pertinent responses to these stimuli are vital for successful social communication [7, 60]. Facial expressions act as easily identifiable signals of the presence of important environmental events [61], allowing quick diffusion of valence information to conspecifics regarding novel stimuli or environments [4]. The changes seen in the fusiform gyrus may also reflect the importance of facial recognition [62]. Ordinarily, it is more difficult to recognize emotions from prosody alone than from facial expressions [63]. Second, there is an emotional processing advantage of pictorial materials over verbal or vocal materials. It has been indicated that pictures have faster and more direct access to meaning than verbal materials [64]. Pictorial stimuli, due to more detailed and vivid description of emotional meanings, often show a processing advantage for evoking a larger or faster emotional effect than verbal materials [65, 66]. By contrast, no consistent processing advantage has been reported for vocal prosodies [28, 67]. A recent study by Cornew *et al.* reported no processing bias for angry prosody but instead found a recognition advantage for neutral over emotional prosodies [68]. The second explanation is also supported by our finding of similar brain activations for facial and bimodal emotions both in the whole brain and in the ROI analysis. That is, neural activations elicited by facial materials were not significantly intensified by the addition of a simultaneous vocal emotion, either in the inferior temporal-occipital activations or in the STS.

However, there is evidence suggesting that the inferior temporal gyrus and its adjacent occipital areas are predominantly involved in emotional face processing [45, 46]. In this regard, it is not surprising to find higher activations during facial than vocal emotion perception. Prior studies have reported involvement of the superior temporal sulcus both in speech-voice processing [69, 70] and in face processing [71, 72], implicating this area in an unbiased, modality-irrelevant processing of emotion. Recently, Peelen *et al.* manipulated the sensory modality of emotional display to investigate whether there are general neural substrates subserving the supramodal representation of perceived emotions [3]. Specifically, this work required participants to rate the intensity of an emotional expression (anger, disgust, fear, happiness, sadness) delivered by diverse modalities (body, face, voice). The results of multivoxel pattern analysis implicated the medial prefrontal cortex and superior temporal sulcus in the modality-independent representation of perceived emotions. Based on these studies, we selected the MPFC and STS as ROIs to examine how the emotional impact of perceiving a specific emotional display varies as a function of modality in these regions, as they are not unique to any modality in terms of emotional processing [3, 73]. We found higher activity for facial than vocal anger expression, but similar activity for facial and bimodal anger expression, in STS activity. These modality effects were not found for happy and neutral expressions.

One explanation for the results for STS activity is emotional negativity bias. Due to our significant adaptive and evolutionary values, the human brain is notably sensitive to emotionally negative events and prioritizes the processing of these events over neutral and positive events [55, 56, 74–76]. This processing bias has also been reported in STS activity for both visual and auditory modalities [28, 45]. For instance, Engell and Haxby (2007) reported enhanced bilateral superior temporal sulcus activations for negative relative to neutral facial expressions, irrespective of the category of the facial expression (fear, disgust, or sadness) [45]. In addition, using two consecutive fMRI experiments, Grandjean *et al.* showed stronger hemodynamic responses in the STS region to angry prosody than to neutral prosody [28]. Due to the emotional negativity bias and the sensitivity of the STS to detecting this bias, the modality effect of emotion perception would be more readily detected by the STS during anger expression, the timely decoding of which is of greater adaptive significance than that of a neutral or a happy expression [77]. This explanation is also supported by the emotion rating data of Study 2. That is, facial relative to prosodic priming had a greater influence on the ratings of bimodal targets only when the primer was an angry expression. Specifically, the facial relative to the prosodic priming of anger intensified the emotion perception for a bimodal angry target while hampering the perception for a bimodal happy target. However, there were no similar effects during the priming of happiness.

However, the MPFC showed similar PSCs for facial, vocal, and bimodal conditions. This may be related to the domain-general functional characteristics of the MPFC, which is considered to support a general function of attention mobilization for mental state decoding [78]. Ochsner *et al.* asked participants to assess the emotional state of themselves or other individuals [79]. The results showed significant MPFC activation during both emotional judgment conditions, regardless of whether the judgement was oriented to self or other [79]. This domain-general function is also evidenced by the involvement of the MPFC in the discrimination of emotion category but not emotion intensity [3, 80]. Though a couple of studies have shown that the MPFC encodes the category of emotional expression well [80, 81], studies using two methods of neuroimaging (positron emission tomography and fMRI) indicate that the MPFC is insensitive to the perceived intensity of emotional display [82], such that emotional expression differing in perceived intensity elicits similar activation amplitudes in this region [3, 82, 83].

Taken together, using both behavioral and neuroimaging approaches, our studies consistently suggested that facial expression elicits more intense emotion perception than the equivalent prosodic expression, despite controlling the emotion intensity for visual and vocal materials. This has implications for emotion regulation by modality-choice in life settings. The process model of emotion regulation proposes that such regulation may be realized by the selection of different situations [84]. However, currently empirical evidence is still lacking concerning how situation selection is performed to realize the goal of emotion regulation. Our results suggest that selecting the vocal modality to receive emotional information is associated with a smaller emotional impact than receiving the same information by the facial modality. That is, imagining that you have trouble with somebody and he/she may lose his/her temper, one may choose to deal with this emotional issue by vocal (e.g. by phone) rather than by face-to-face contact to reduce emotional stress. Alternatively, when you want others to be happy, you can just put a big smile on your face and let them see it.

Several limitations and future directions need to be noted. First, there was a fixation in the visual modality, while there was no similar cue in the vocal modality. This may cause cross-modal imbalance in the initial occupation of attention, so bimodal fixation needs to be designed in future studies. Second, although the results of Study 2 confirmed the conclusion of Study 1 using the behavioral approach, it remains undetermined whether the activation for the facial *versus* prosodic contrast in the temporal-occipital area is independent of their roles in vision-specific processing. Third, we did not have time-jittering between the neutral and emotional expressions to distinguish them. Fourth, although our study focused on the modality effect in emotional perception between facial expression and vocal prosodic materials, the lack of differences between the unimodal and bimodal conditions cannot be completely explained. In further studies, we aim to address this issue, control the bimodal materials, and explain these results in depth. Moreover, we have to confine the current conclusions to angry and happy expressions. That is, future studies are needed to determine whether the current findings can be generalized to other categories of emotional expression like fear, sadness, or disgust. Last, it is known that emotional expression elicits not only emotional arousal but also emotion recognition [85]. In this regard, future studies are needed to replicate the current findings using emotionally evocative scenes and sounds, which are more specific to emotional arousal [85–87], in order to elucidate the practical implications of the current findings for emotion regulation.

Acknowledgements This work was supported by the National Natural Science Foundation of China (31371042 and 31671164). We are greatly indebted to Siyu Zhu and Ling Zhang for their support. We acknowledge the invaluable help of Xiang Li in supporting the fMRI data processing.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Keltner D, Haidt J. Social functions of emotions at four levels of analysis. Cogn Emotion 1999, 13: 505–521.
2. Oatley K, Keltner D, Jenkins JM. Understanding Emotions (2nd ed.). Netherlands: Springer, 2006.
3. Peelen MV, Atkinson AP, Vuilleumier P. Supramodal representations of perceived emotions in the human brain. J Neurosci 2010, 30: 10127–10134.
4. Blair RJR, Morris JS, Frith CD, Perrett DI, Dolan RJ. Dissociable neural responses to facial expressions of sadness and anger. Brain 1999, 122 (Pt 5): 883.
5. Ethofer T, Kreifelts B, Wiethoff S, Wolf J, Grodd W, Vuilleumier P*, et al.* Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. J Cogn Neurosci 2009, 21: 1255.
6. Heberlein AS, Atkinson AP. Neuroscientific evidence for simulation and shared substrates in emotion recognition. Emotion Rev 2009, 1: 162–177.
7. Surguladze S, Brammer MJ, Keedwell P, Giampietro V, Young AW, Travis MJ*, et al.* A differential pattern of neural response toward sad versus happy facial expressions in major depressive disorder. Biol Psychiat 2005, 57: 201–209.
8. Adolphs R, Tranel D, Damasio H, Damasio A. Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature 1994, 372: 669–672.
9. Adolphs R, Tranel D, Damasio H, Damasio AR. Fear and the human amygdala. J Neurosci 1995, 15: 5879–5891.
10. Calder AJ, Young AW, Rowland D, Perrett DI, Hodges JR, Etcoff NL. Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. Cogn Neuropsychol 1996, 13: 699–745.
11. Young AW, Aggleton JP, Hellawell DJ, Johnson M, Broks P, Hanley JR. Face processing impairments after amygdalotomy. Brain 1995, 118: 15.
12. Ojemann JG, Ojemann GA, Lettich E. Neuronal activity related to faces and matching in human right nondominant temporal cortex. Brain 1992, 115 Pt 1: 1.
13. Rapcsak SZ, Comer JF, Rubens AB. Anomia for facial expressions: Neuropsychological mechanisms and anatomical correlates. Brain Lan 1993, 45: 233–252.
14. Cancelliere AE, Kertesz A. Lesion localization in acquired deficits of emotional expression and comprehension. Brain Cogn 1990, 13: 133–147.
15. Adolphs R, Damasio H, Tranel D, Damasio AR. Cortical systems for the recognition of emotion in facial expressions. J Neurosci 1996, 16: 7678–7687.
16. Imaizumi S, Mori K, Kiritani S, Hosoi H, Tonoike M. Task-dependent laterality for cue decoding during spoken language processing. Neuro Rep 1998, 9: 899–903.
17. Ross ED. Affective prosody and the aprosodias. In Mesulam MM (Ed.), Principles of Behavioral and Cognitive Neurology. New York, NY: Oxford University Press, 2000: 316–331.
18. Mitchell RL, Crow TJ. Right hemisphere language functions and schizophrenia: the forgotten hemisphere? Brain 2005, 128: 963–978.
19. Mitchell RL, Elliott R, Barry M, Cruttenden A, Woodruff PW. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. Neuropsychologia 2003, 41: 1410.
20. George MS, Parekh PI, Rosinsky N, Ketter TA, Kimbrell TA, Heilman KM*, et al.* Understanding emotional prosody activates right hemisphere regions. Arch of Neurol 1996, 53: 665–670.
21. Buchanan TW, Lutz K, Mirzazade S, Specht K, Shah NJ, Zilles K*, et al.* Recognition of emotional prosody and verbal components of spoken language: An fMRI study. Cogn Brain Res 2000, 9: 227–238.
22. Gur RC, Skolnick BE, Gur RE. Effects of emotional discrimination tasks on cerebral blood flow: Regional activation and its relation to performance. Brain Cogn 1994, 25: 271–286.
23. Breiter HC, Etcoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL*, et al.* Response and habituation of the human amygdala during visual processing of facial expression. Neuron 1996, 17: 875–887.
24. Phillips ML, Bullmore ET, Howard R, Woodruff PW, Wright IC, Williams SC*, et al.* Investigation of facial recognition memory and happy and sad facial expression perception: an fMRI study. Psychiat Res 1998, 83: 127–138.
25. Keslerwest ML, Andersen AH, Smith CD, Avison MJ, Davis CE, Kryscio RJ*, et al.* Neural substrates of facial emotion processing using fMRI. Brain Res Cogn Brain Res 2001, 11: 213–226.
26. Sprengelmeyer R, Rausch M, Eysel UT, Przuntek H. Neural structures associated with recognition of facial expressions of basic emotions. Proc Biol Sci 1998, 265: 1927.
27. Johnstone T, Reekum CMV, Oakes TR, Davidson RJ. The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions. Soc Cogn Affect Neurosci 2006, 1: 242–249.
28. Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR*, et al.* The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci 2005, 8: 145.
29. Koelsch S, Fritz T, Dy VC, Müller K, Friederici AD. Investigating emotion with music: an fMRI study. Hum Brain Mapp 2006, 27: 239–250.
30. Hulka LM, Preller KH, Matthias V, Broicher SD, Quednow BB. Cocaine users manifest impaired prosodic and cross-modal emotion processing. Front Psychiat 2013, 4: 98.
31. Chen X, Pan Z, Wang P, Yang X, Liu P, You X*, et al.* The integration of facial and vocal cues during emotional change perception: EEG markers. Soc Cogn Affect Neurosci 2016, 11: 1152–1161.
32. Biele C, Grabowska A. Sex differences in perception of emotion intensity in dynamic and static facial expressions. Exp Brain Res 2006, 171: 1.
33. Uono S, Sato W, Toichi M. Brief report: Representational momentum for dynamic facial expressions in pervasive developmental disorder. J Autism Dev Dis 2010, 40: 371–377.
34. Yoshikawa S, Sato W. Enhanced perceptual, emotional, and motor processing in response to dynamic facial expressions of emotion1. Jap Psychol Res 2006, 48: 213–222.
35. Kilts CD, Egan G, Gideon DA, Ely TD, Hoffman JM. Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. Neuroimage 2003, 18: 156–168.
36. Labar KS, Crupain MJ, Voyvodic JT, Mccarthy G. Dynamic perception of facial affect and identity in the human brain. Cereb Cortex 2003, 13: 1023–1033.
37. Sato W, Kochiyama T, Yoshikawa S, Naito E, Matsumura M. Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. Cogn Brain Res 2004, 20: 81–91.
38. Trautmann SA, Fehr T, Herrmann M. Emotions in motion: Dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. Brain Res 2009, 1284: 100–115.
39. Gong X, Huang YX, Wang Y, Luo YJ. Revision of the Chinese facial affective picture system. Chin Ment Health J 2011, 25: 40–46.
40. Chen X, Zhao L, Jiang A, Yang Y. Event-related potential correlates of the expectancy violation effect during emotional prosody processing. Biol Psychol 2011, 86: 158–167.

41. Chen X, Yang J, Gan S, Yang Y. The contribution of sound intensity in vocal emotion perception: Behavioral and electrophysiological evidence. Plos One 2012, 7: e30278.
42. Kotz SA, Paulmann S. When emotional prosody and semantics dance cheek to cheek: ERP evidence. Brain Res 2007, 1151: 107–118.
43. Yan C, Zang Y. DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. Front Syst Neurosci 2010, 4: 13.
44. Brett M, Anton JL, Valabregue R, Poline JB. Region of interest analysis an SPM toolbox. Neuroimage 2002, 16.
45. Engell AD, Haxby JV. Facial expression and gaze-direction in human superior temporal sulcus. Neuropsychologia 2007, 45: 3234–3241.
46. Vuilleumier P, Pourtois G. Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. Neuropsychologia 2007, 45: 174–194.
47. Stigliani A, Jeska B, Grill-Spector K. Encoding model of temporal processing in human visual cortex. Proc Nat Acad Sci U S A 2017, 114: E11047.
48. Grillspector K, Kushnir T, Hendler T, Edelman S, Itzchak Y, Malach R. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum Brain Mapp 2015, 6: 316–328.
49. Adolphs R. Recognizing emotion from facial expressions: Psychological and neurological mechanisms. Behav Cogn Neurosc Rev 2002, 1: 21.
50. Leppanen JM, Hietanen JK. Affect and face perception: odors modulate the recognition advantage of happy faces. Emotion 2003, 3: 315–326.
51. Hugdahl K, Iversen PM, Johnsen BH. Laterality for facial expressions: Does the sex of the subject interact with the sex of the stimulus face? Cortex 1993, 29: 325–331.
52. Kirita T, Endo M. Happy face advantage in recognizing facial expressions. Acta Psychologica 1995, 89: 149–163.
53. Farran EK, Branson A, King BJ. Visual search for basic emotional expressions in autism; impaired processing of anger, fear and sadness, but a typical happy face advantage. Res Autism Spect Dis 2011, 5: 455–462.
54. Leppänen JM, Kauppinen P, Peltola MJ, Hietanen JK. Differential electrocortical responses to increasing intensities of fearful and happy emotional expressions. Brain Res 2007, 1166: 103–109.
55. Yuan JJ, Li H, Chen AT, Luo YJ. Neural correlates underlying humans' differential sensitivity to emotionally negative stimuli of varying valences: an ERP study. Prog Nat Sci 2007, 32: 115–121.
56. Yuan J, Zhang Q, Chen A, Li H, Wang Q, Zhuang Z, *et al.* Are we sensitive to valence differences in emotionally negative stimuli? Electrophysiological evidence from an ERP study. Neuropsychologia 2007, 45: 2764.
57. Sprengelmeyer R, Jentzsch I. Event related potentials and the perception of intensity in facial expressions. Neuropsychologia 2006, 44: 2899–2906.
58. Leppänen JM, Hietanen JK. Positive facial expressions are recognized faster than negative facial expressions, but why? Psychol Res 2004, 69: 22–29.
59. Wild B, Erb M, Eyb M, Bartels M, Grodd W. Why are smiles contagious? An fMRI study of the interaction between perception of facial affect and facial movements. Psychiat Res 2003, 123: 17–36.
60. Darwin C. The Expression of the Emotions in Man and Animals (3rd ed.). New York: Oxford University Press, 1872.
61. Surguladze SA, Brammer MJ, Young AW, Andrew C, Travis MJ, Williams SCR, *et al.* A preferential increase in the extrastriate response to signals of danger. Neuroimage 2003, 19: 1317–1328.
62. Fan L, Jiang T. Mapping underlying maturational changes in human brain. Neurosci Bull 2017: 1–3.
63. Adolphs R. Neural systems for recognizing emotion. Curr Opin Neurobiol 2002, 12: 169–177.
64. Azizian A, Watson TD, Parvaz MA, Squires NK. Time course of processes underlying picture and word evaluation: An event-related potential approach. Brain Topogr 2006, 18: 213–222.
65. Kensinger EA, Schacter DL. Processing emotional pictures and words: Effects of valence and arousal. Cogn Affect Behav Neurosci 2006, 6: 110–126.
66. Schacht A, Sommer W. Emotions in word and face processing: Early and late cortical responses. Brain Cogn 2009, 69: 538–550.
67. Kai A, Rank E, Kotz SA, Toepel U, Besson M, Schirmer A, *et al.* Affective encoding in the speech signal and in event-related brain potentials. Speech Commun 2003, 40: 61–70.
68. Cornew L, Carver LJ, Love T. There's more to emotion than meets the eye: A processing bias for neutral content in the domain of emotional prosody. Cogn Emotion 2010, 24: 1133–1152.
69. Mottonen R, Calvert GA, Jaaskelainen IP, Matthews PM, Thesen T, Tuomainen J, *et al.* Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. Neuroimage 2006, 30: 563–569.
70. Uppenkamp S, Johnsrude IS, Norris D, Marslenwilson WD, Patterson RD. Locating the initial stages of speech-sound processing in human temporal cortex. Neuroimage 2006, 31: 1284–1296.
71. Ishai A, Schmidt CF, Boesiger P. Face perception is mediated by a distributed cortical network. Brain Res Bull 2005, 67: 87–93.
72. Tseng LY, Tseng P, Liang W, Hung DL, Tzeng OJL, Muggleton NG, *et al.* The role of superior temporal sulcus in the control of irrelevant emotional face processing: A transcranial direct current stimulation study. Neuropsychologia 2014, 64: 124–133.
73. Hein G, Knight RT. Superior temporal sulcus—it's my area: Or is it? J Cogn Neurosci 2008, 20: 2125–2136.
74. Delplanque S, Lavoie ME, Hot P, Silvert L, Sequeira H. Modulation of cognitive processing by emotional valence studied through event-related potentials in humans. Neurosci Lett 2004, 356: 1–4.
75. Delplanque S, Silvert L, Hot P, Rigoulot S, Sequeira H. Arousal and valence effects on event-related P3a and P3b during emotional categorization. Int J Psychophysiol 2006, 60: 315–322.
76. Huang YX, Luo YJ. Temporal course of emotional negativity bias: an ERP study. Neurosci Lett 2006, 398: 91–96.
77. Horstmann G, Bauland A. Search asymmetries with real faces: Testing the anger-superiority effect. Emotion 2006, 6: 193–207.
78. Gallagher HL, Frith CD. Functional imaging of 'theory of mind'. Trends Cogn Sci 2003, 7: 77–83.
79. Ochsner KN, Knierim K, Ludlow D, Hanelin J, Ramachandran T, Glover GH, *et al.* Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. J Cogn Neurosci 2004, 16: 1746–1772.
80. Kesler-West ML, Andersen AH, Smith CD, Avison MJ, Davis CE, Kryscio RJ, *et al.* Neural substrates of facial emotion processing using fMRI. Cogn Brain Res 2001, 11: 213–226.
81. Phillips ML, Bullmore ET, Howard R, Woodruff PWR, Wright IC, Williams SCR, *et al.* Investigation of facial recognition memory and happy and sad facial expression perception: an fMRI study. Psychiatry Res 1998, 83: 127–138.
82. Ndiaye KBJ, Sander D, Vuilleumier P. Self-relevance processing in the human amygdala: gaze direction, facial expression, and emotion intensity. Emotion 2009, 9: 798–806.
83. Phillips ML, Young AW, Senior C, Brammer M, Andrew C, Calder AJ, *et al.* A specific neural substrate for perceiving facial expressions of disgust. Nature 1997, 389: 495.

84. Gross JJ, Thompson RA. Emotion regulation: Conceptual foundations. In Gross JJ (Ed.). Handbook of Emotion Regulation. New York, NY: Guilford Press, 2007: 3–24.
85. Britton JC, Phan KL, Taylor SF, Welsh RC, Berridge KC, Liberzon I. Neural correlates of social and nonsocial emotions: An fMRI study. Neuroimage 2006, 31: 397–409.
86. Bradley MM, Lang PJ. Affective reactions to acoustic stimuli. Psychophysiology 2000, 37: 204–215.
87. Yuan J, Xu S, Yang J, Liu Q, Chen A, Zhu L*, et al.* Pleasant mood intensifies brain processing of cognitive control: ERP correlates. Biol Psychol 2011, 87: 17–24.