NeuroImage 203 (2019) 116199

Contents lists available at ScienceDirect

NeuroImage

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

Modality-general representations of valences perceived from visual and auditory modalities

Jin Gu[a,1], Linjing Cao[a,1], Baolin Liu[b,*]

[a] College of Intelligence and Computing, Tianjin University, Tianjin, 300350, PR China
[b] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, 100083, PR China

A R T I C L E I N F O

Keywords:
Valence
Modality-general representation
Left postcentral gyrus
MTG
Visual
Auditory

A B S T R A C T

Valence is a dimension of emotion and can be either positive, negative, or neutral. Valences can be expressed through the visual and auditory modalities, and the valences of each modality can be conveyed by different types of stimuli (face, body, voice or music). This study focused on the modality-general representations of valences, that is, valence information can be shared across not only visual and auditory modalities but also different types of stimuli within each modality. Functional magnetic resonance imaging (fMRI) data were collected when subjects made affective judgment on silent videos (face and body) and audio clips (voice and music). The searchlight analysis helped to locate four areas that might be sensitive to the representations of modality-general valences, including the bilateral postcentral gyrus, left middle temporal gyrus (MTG) and right middle frontal gyrus (MFG). Further cross-modal classification based on multivoxel pattern analysis (MVPA) was performed as a validation analysis, which suggested that only the left postcentral gyrus could successfully distinguish three valences (positive versus negative and versus neutral: PvsNvs0) across different types of stimuli (face, body, voice or music), and the classification was also successful in left MTG across the stimuli types of face and body. The univariate analysis further found the valence-specific activation differences across stimulus types in MTG. Our study showed that the left postcentral gyrus was informative to valence representations, and extended the research about valence representation that the modality-general representation of valences across not only visual and auditory modalities but also different types of stimuli within each modality.

1. Introduction

Emotional communication with others plays an important role in social interaction. In particular, emotional expressions can provide key information about the emotional state of the presenter or the surrounding environment. Therefore, these signals can help guide our behavior and even have survival value in some cases. Affective information can be transmitted through different modalities and channels, such as vocal intonations, facial expressions or body movements. In addition, music is also a powerful means of expressing emotions and there were brain regions characterizing the emotions of music (Koelsch et al., 2006). Although the affective information can be expressed in different modalities, these signals can similarly lead to the recognition of emotional states. Therefore, affective information might be coded strictly across different modalities, which we call the modality-general representations of emotions.

In addition to studies on categorical emotions, the dimensional methods of emotion can provide different interpretations of affective processing. The categorical method emphasizes different neural circuits for specific emotions (e.g. disgust, sadness, and happiness), while the dimensional method advocates the neural representation of potential dimensions including valence, arousal and other higher-level dimensions (Barrett and Russell, 2014; Fontaine et al., 2007). Recent studies using functional magnetic resonance imaging (fMRI) to examine the neural representation of categorical emotions caused by multiple stimulus types have provided evidence for modality-general processing (Kim et al., 2015; Peelen et al., 2010). Peelen et al. (2010) found that five emotions (disgust, anger, fear, sadness, and happiness) can be represented across three emotional stimuli (face, body, and voice), which provided evidence of modality-general representations of emotions. They found that the common affective processing of the brain is located in the medial prefrontal cortex (mPFC) and the left superior sulcus (STS). In addition, the

* Corresponding author.
E-mail address: liubaolin@tsinghua.edu.cn (B. Liu).
[1] Jin Gu and Linjing Cao contribute equally to this work.

https://doi.org/10.1016/j.neuroimage.2019.116199
Received 26 November 2018; Received in revised form 31 August 2019; Accepted 14 September 2019
Available online 16 September 2019
1053-8119/

modality-general representations were shown to be induced by the memory of emotional events (Kim et al., 2015). Some studies using multivoxel pattern analysis (MVPA) also found modality-general representations of basic emotions (Kassam et al., 2013; Klasen et al., 2011; Kragel and LaBar, 2015; Saarimaki et al., 2016). A study also found that there is a significant cross-modal adaptation of effect between voice and face stimuli, which supports the representations of modality-general emotions (Watson et al., 2014). Our latest research has found that emotions can be represented in left postcentral gyrus independent from different stimulus types in the visual modality (face, body, and whole-person) (Cao et al., 2018).

Previous studies mostly focus on the specific emotions, and only recently few studies began to focus on the valence of affects (positive, negative and neutral) represented across different modalities. Chikazoe et al. and Kim et al. (Chikazoe et al., 2014; Kim et al., 2017) have found a general representation of the valence across modalities. Chikazoe et al. presented gustatory stimuli and pictures to participants and found modality-general representations in the lateral and medial frontal cortex. Kim et al. conducted a neuroimaging study to examine valence encoding across visual and auditory modalities using music clips and silent videos and found the common neural representations of valence. Shinkareva et al. conducted a neuroimaging study that examined valence processing across modalities using visual images and auditory sounds, both of which can elicit emotional responses (Shinkareva et al., 2014). However, the results provided evidence supporting modality-specific valence representations but not modality-general representations. Kim et al. conducted a searchlight analysis of audio-visual emotional stimuli with different valence and arousal (Kim et al., 2016). This study located five brain regions that represent valence: the right posterior cingulate cortex, the left mPFC, the left superior/middle temporal gyrus, the MFG, and the thalamus. However, it is limited to the use of audiovisual stimuli in the natural situation. It is not clear whether the abstract representation of emotional valence is independent of all single modal (visual or auditory) stimuli.

Above all, the study about the characteristics of modality-general representations of emotional valence deserves more attention. In addition, it is still unclear whether the representation of valence could be shared across not only visual and auditory modalities but also different types of stimuli within each modality. The current study investigated the potential role of different stimulus types in the modality-general representation of valence by including two different stimulus types in each modality (visual or auditory modality). The visual stimuli contained emotional face and body, and the auditory stimuli contained affective voice and music. The main purpose of our research is to use fMRI to provide additional evidence for the general encoding of valence expressed through different types of stimuli within visual and auditory modalities.

In the present work, we studied the valence representations across stimulus types (face, body, voice, and music), respectively from the visual and auditory modalities. fMRI data were acquired when thirty-one participants made valence judgment on the stimuli. We first used searchlight analysis (Kriegeskorte et al., 2006) to locate brain regions showing modality-general valences regardless of different modalities (visual and auditory) and different stimulus types within each modality. Then, we performed cross-modal classification and univariate analysis as additional validations. In the cross-modal classification of valence, the classifier was trained from data of three modalities (e.g. body or voice or music) and then was used to predict valence categories of data from the remaining modality (e.g. face). The classification tested whether the valences can be successfully distinguished across different modalities and stimulus types in the clusters located by searchlight analysis. Finally, a further univariate analysis was performed to explore the differences between the average activation of different conditions in the regions where valences can be classified successfully. The results showed the modality-general representations of valence in the left postcentral gyrus.

2. Materials and methods

2.1. Participants

Thirty-one healthy volunteers participated in this study. All subjects were right-handed without any hearing impairment, and with normal or corrected-to-normal vision, and all of them had no history of neurological or psychiatric illness. Five subjects were excluded due to larger head movements, and the data of twenty-six subjects were eventually included for further analysis (fourteen females, average age 21.8 ± 1.8 years, range from 19 to 26 years old). This study was approved by the Institutional Review Board of Binzhou Medical University and written informed consent was obtained from all subjects before the experiments. Another group of adult volunteers ($n = 18$) participated in the preliminary behavioral experiments to select the most effective stimuli in terms of valence and arousal.

2.2. Experiment stimuli

Silent video clips of emotional body expressions and face expressions and the emotional voice stimuli were chosen from the Geneva Multimodal Emotion Portrayals (GEMEP) corpus (Banziger et al., 2012). The videos were cropped by removing and masking irrelevant aspects such that non-facial body parts or facial features were not visible in video clips of face or body. The silent video clips and voices expressed three different valences (negative, neutral, and positive) as the previous studies (Kim et al., 2016, 2017). In addition, the positive and negative valences included both high and low arousal degrees in our study. Four actors (2 male) expressed each valence category. We checked the quantitative difference in the amount of movement in the face and body videos, and the processing method of the face and body clips was the same as that in our previous research (Cao et al., 2018). For details of the processing operations, please refer to our previous paper, it would not be elaborated in this paper. One-way ANOVA showed that there was no significant difference in the movement estimation of different valences ($F = 1.08$, $p = 0.35$). The emotional music stimuli were selected from the music set which was created and validated by Eerola et al. (Eerola and Vuoskoski, 2010) and used in Kim et al. (2017). All videos, voices as well as music clips were clipped or combined to exactly 2s by editing longer or shorter clips separately. The representative stimuli in visual modality were presented in Fig. 1A.

Prior to the current study, another group of volunteers (9 female, mean age: 22.3 years; 9 males, mean age: 21.9 years) performed a separate behavioral norming experiment. These subjects did not have any known difficulties when performing emotional recognition. They were asked to classify the emotional stimuli with three labels (positive, negative, and neutral) and rate the emotional arousal for the positive and negative stimuli according to a 9-point scale (1 lowest arousal, 9 highest arousal). In order to ensure there is a significant difference between low–arousal and high-arousal stimuli. We selected stimuli with scores less than 4 and scores greater than 6.5 points in the arousal dimension. At last, the stimuli used in our experiment consisted of a total of 40 video clips and 40 audio clips (5 emotion conditions $\times$ 4 stimulus types $\times$ 4 exemplars per condition). The five emotion conditions include low-arousal positive valence, high-arousal positive valence, neutral valence, low-arousal negative valence, and high-arousal negative valence. In the stimulus materials, positive emotion included happiness and the negative emotion include anger and fear. In the current study, the stimuli were collapsed across arousal levels and this dimension would not be considered in the current manuscript. Therefore, both the high and low arousal blocks were used as a whole according to the valence (positive or negative). All in all, in the current study there are 12 stimuli conditions (3 valences $\times$ 4 stimulus types). All videos, voices, and music clips were selected to effectively express three valence categories, and all expressions were well-recognized (accuracy $>$ 83%).

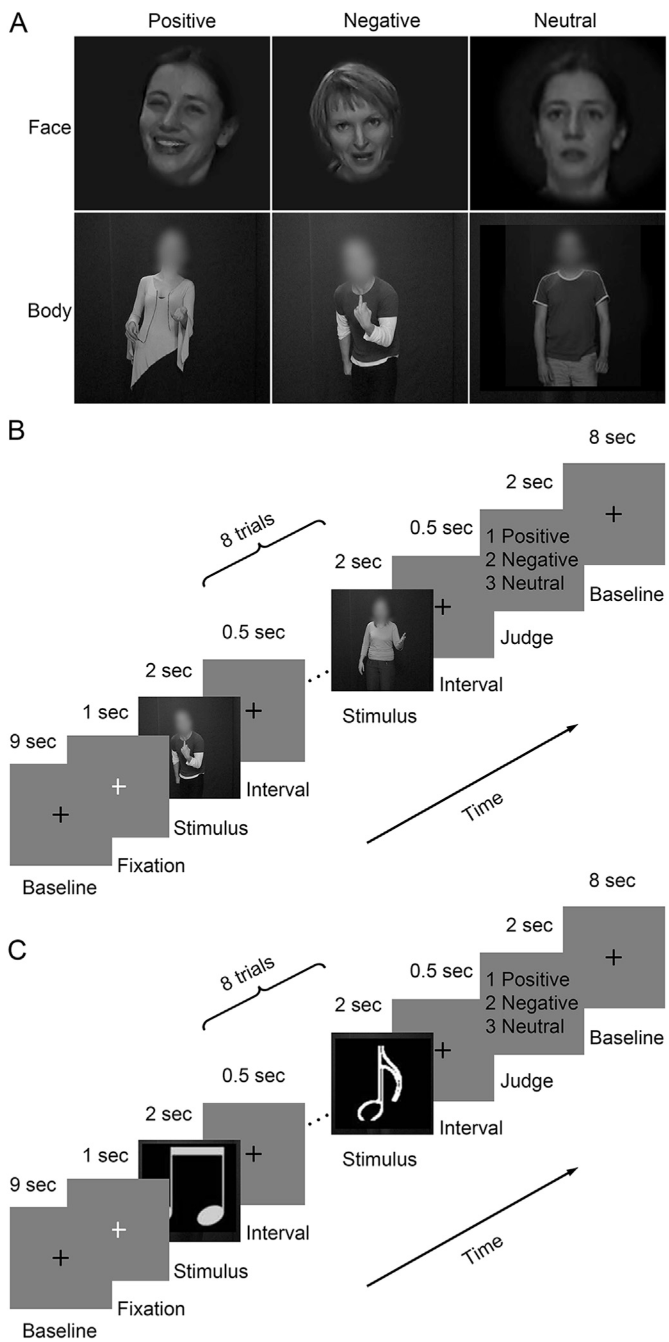

Fig. 1. The representative stimuli in the visual modality and paradigm of the experiment. (A) Two types of stimuli in the visual modality (face and body) expressing three valence categories (positive, negative and neutral) were used in the experiment. (B–C) A brief overview of the presentation timing of the valence recognition task. The subjects performed five runs with a fixed baseline period of 10 s at the beginning and end of each run. In each run, there are 24 pseudo-randomly blocks with a fixed interval of 8s between the blocks. Each block contains 8 trials, which presents visual stimuli (face or body) or auditory stimuli (voice or music) of the same valence category (positive or negative or neutral). Each stimulus is presented for 2 s in each trial, then a 0.5s interval. Then at the end of each block, subjects were asked to make choices among the three valences within 2 s, while ignoring the stimulus types that exhibited the valence.

2.3. Procedure

The fMRI procedure consists of five runs. Since the positive and negative valences include both high and low arousal degrees, the blocks of the neutral stimuli are repeated twice to ensure that the data amounts of the three valence categories are equal in each run. Therefore there are 24 pseudo-randomly blocks (2 valences × 2 arousal degrees × 4 stimuli categories + neutral valence × 4 stimuli categories × repeat 2 times) with a fixed interval of 8s between the blocks. Each block contains 8 trials, which presented stimuli from the same stimulus type (face or body or voice or music) and the same valence (positive or negative or neutral). Each stimulus is presented for 2 s with a 0.5s inter-stimulus interval. At the end of each block, subjects were asked to recognize the emotion valence shown in the stimuli within 2s. As noted, stimuli might inevitably involve memory or other emotion-independent cognitive domains driving the similarities in activation maps, which is a limitation in the experiments about advanced cognition. Participants were instructed to focus on the valence of stimuli and the valence judgment task also helped participants maintain attention which helped us minimize the impact as much as possible. Fig. 1B is a schematic diagram of the visual block and Fig. 1C is a schematic diagram of the auditory block.

2.4. Data acquisition

Earplugs and foam pads were used to reduce the scanner noise and head movement. First we used the echo planar image sequence (EPI) to obtain the T2 *weighted images (TR = 2000 ms, TE = 30 ms, voxel size = 3.1 × 3.1 × 4.0 mm^3, matrix size = 64 × 64, slices = 33, slices thickness = 4 mm, slices gap = 0.6 mm, FA = 90°) for each subject. We then acquired the high-resolution anatomical localization images using a three-dimensional magnetization-prepared rapid-acquisition gradient echo sequence (TR = 1900 ms, TI = 1100 ms, TE = 2.52 ms, matrix size = 256 × 256, voxel size = 1 × 1 × 1 mm^3, FA = 90°) at the end of each scan. The stimuli were displayed to participants by the VisuaStim digital (Resonance Technology Company, Inc.), which is equipped with a corresponding headset and glasses. The sound stimuli were displayed by electrostatic headphones and the visual stimuli were presented through high-resolution stereo 3D glasses.

2.5. Data analysis

2.5.1. Behavioral measures

The classification accuracy and response time of three valences in four stimulus types were recorded for each participant. Then the SPSS 18 software was used to perform the repeated measures ANOVA for the response time with the factors of valence (positive, negative, and neutral) and stimulus types (face, body, voice, and music) to test the interactions and main effects. In addition, we performed the follow-up tests for the significant interaction effect to test all stimulus types against each other within a specific valence or the all valences against each other within a specific modality by separate ANOVA.

2.5.2. Data preprocessing

We used the SPM8 software package (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) which integrated into Matlab software (The Math Works) to preprocess and analyze the functional images. First, we discarded the volumes collected in the first 10s for each run to allow for the equilibration effects. Then the remaining data of each run were slice-timing corrected to the middle slice and realigned to the first image for head-movement correction. Subsequently, the high-resolution T1-weighted images were co-registered to the mean functional images and were segmented into the gray matter, white matter and cerebrospinal fluid (CSF) for normalization. Then the generated parameters were used to spatially normalize the functional images into the Montreal Neurological Institute (MNI) space and were resampled to 3 × 3 × 3 mm^3. Finally, the functional data in the five runs were smoothed with a 4-mm full-width half-maximum (FWHM) Gaussian kernel (Liang et al., 2018; Zhang et al., 2016). The fMRI data were further analyzed using the general linear model (GLM) to construct the regressors for each 12 experimental condition (positive face, negative face, neutral face, positive body, negative body, neutral body, positive voice, negative voice and neutral voice, positive music, negative music, neutral music). In each of the contrast, one of the 12 stimulus conditions was set as 1 and others were set as 0, therefore we extracted the neural activity of the current condition, as shown in Fig. 2A. The regression models were estimated for different conditions. We applied topographical FDR correction with the threshold of $q = 0.05$ based on an uncorrected voxel-wise threshold of $p = 0.001$ in SPM8. All the above data preprocessing was conducted on each subject.

2.5.3. Searchlight analysis

To localize regions that may respond to valences across stimuli types (body or face or voice or music), searchlight analysis was performed (Kriegeskorte et al., 2006). First, the neural activity has been extracted for each of the 12 stimulus conditions (3 valences × 4 stimulus types) as shown in the section of data preprocessing. Then, a sphere of 6 mm radius centered on each voxel was defined and the activities of the neighboring voxels within the radius were considered as a whole, which was used in the next searchlight processing. As illustrated in Fig. 2B, pair-wise correlations were then computed between any two different activity patterns in the 12 stimulus conditions using the correlation distance (Pearson correlation), resulting in a symmetrical 12 × 12 correlation matrix for the sphere of 6 mm radius, which was considered as the result of the voxel at the center. Therefore, we can get a 12x12 matrix of correlations for the voxel. Then this procedure was repeated at each voxel across the whole brain so that we obtained 12x12 matrix of correlations for each voxel. This matrix composed of five parts: within-face, within-body, within-voice, within-music, and between-modality correlation coefficients. Peelen et al. (2010) had used a similar technique successfully in their searchlight analysis and indicated that the representations of modality-general emotions can be diagnosed by between-modality correlation coefficients. We focused on the correlation of between-modality, here we meant six pairs: face-music, face-body, face-voice, music-voice, music-body, body-voice. First, we calculated the correlation coefficients of within-valence, which means the pair is from the same valence, for example, the neural face and neural voice, positive face and positive voice, negative face and negative voice, the same is in the rest five pairs (face-music, face-body, music-voice, music-body, body-voice). Then, we also calculated the correlation coefficients of between-valence, means the correlation between four stimulus types in different valence, for example, positive face versus negative body, neutral body versus positive voice. The related definition referred the work of Kim et al. (2017). As illustrated in Fig. 2B, for the between-stimulus types, the correlation coefficients of within-valence and between-valence were averaged separately, and then the average coefficient of between-valence was subtracted from the average coefficient of within-valence. Fisher z-transform has been applied to the average coefficient and then a difference map (d-map) was generated for each subject. Then we performed one-sample t-tests at each voxel on the group level to test whether the coefficient was non-zero. The significant clusters can be located by the threshold of $p < 0.05$ (FDR corrected) (Woo et al., 2014).

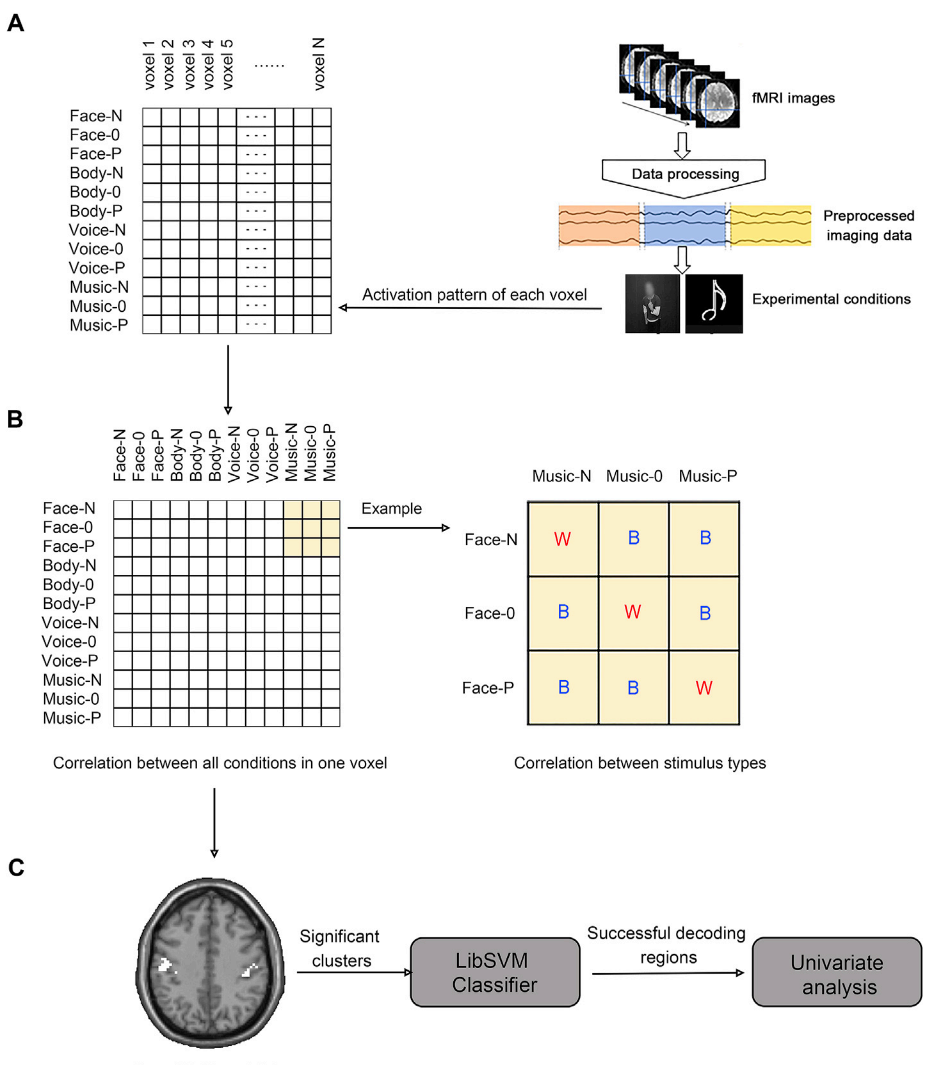

Fig. 2. The main analytical steps of the study. (A) For each of the 12 stimulus conditions (3 valences × 4 stimulus types), the neural activity patterns estimated by GLM were extracted after the fMRI data preprocessing. (B) Pair-wise correlations were computed between any two different activity patterns in the twelve stimulus conditions using the correlation distance (Pearson correlation). Here, we adopted a sphere centered on the voxel and the activities of the neighboring voxels within the radius were considered as a whole. Then pair-wise correlations were computed between any two different activity patterns in the twelve stimulus conditions, resulting in a symmetrical 12 × 12 correlation matrix for the sphere of 6 mm radius, which was considered as the result of the voxel at the center. This matrix was composed of five parts: within-face, within-body, within-voice, within-music, and between-modality correlation coefficients. For the between-modality conditions, the correlation coefficients of within-valence and between-valence were averaged separately, and then the average coefficient of between-valence was subtracted from the average coefficient of within-valence. P = positive, 0 = neutral, N = negative, W = correlation coefficient of within-valence (i.e., positive-face vs positive music, negative-body vs negative-voice, etc.), B = correlation coefficient of between-valence (i.e., positive-face vs negative-music, positive-body vs neutral-voice, etc.). The related definition referred the work of Kim et al. (Kim et al., 2017) (C) The procedure of calculating correlation coefficients was repeated across all the voxels in the searchlight analysis, generating a difference map (d-map) for each subject. Then cross-modal classification was performed as the validation technique at significant clusters located by the searchlight analysis. At last, the univariate analysis was used to test whether there were valence-specific activation differences across modalities in clusters classifying valence successfully.

2.5.4. Cross-modal classification

The valence classification at the cluster-level was performed as the additional validation analysis to examine whether these clusters truly provide information of modality-general valence. The verification is necessary because the significant searchlight clusters were not guaranteed to provide valence information (Etzel et al., 2013). Based on the cross-modal classification, the validation analysis can test whether these clusters carry enough information to represent the modality-general valence across the stimulus types. The analysis can furtherly clarify the specific properties of emotional modality-general valence representation from a different perspective.

The cross-modal classification used in the current study was similar to the classification used in our previous study (Cao et al., 2018) and other exploration of emotional space (Baucom et al., 2012; Kim et al., 2017; Shinkareva et al., 2014). The SVM classifier based on multi-voxel pattern analysis (MVPA) was used for cross-modal classification to examine the modality-general representations of valence at the group level. We conducted a three-way classification of the valences (PvsNvs0). In each cluster identified in searchlight analysis, we extracted the activation pattern under the 12 stimulus conditions by MVPA as the feature data, the three-way classifier was trained by the data from three stimulus types (i.e., body, voice, and music) and then the data from the remaining one (i.e., face) were used as a test set. We used the LibSVM toolkit by Chang et al. (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) which was integrated into Matlab to preprocess the classification (Chang et al., 2011). In order to obtain stable estimates for individual-level accuracy, we have run 1000 repetitions classification and adopted the average results. The classification process was performed in the four types and then the accuracies were averaged among the four cross-validations. For the classification analysis of the three valences, a one-sample *t*-test was performed to assess whether the group average accuracy was significantly higher than the chance level (0.33).

2.5.5. Analysis of the differences between mean activation

We used univariate analysis to explore the activation of the clusters identified from the cross-modal classification under the different conditions (3 valences × 4 stimulus types). As a validation analysis, there is a limitation of double-dipping in the validation due to the using of the same data for selection and selective analysis) (Etzel et al., 2013). To ensure statistical independence of the results, the univariate analysis used the average beta values as the dataset (Kriegeskorte et al., 2009). The beta values for each condition were extracted and averaged across voxels in the cluster (Kim et al., 2015; Peelen et al., 2010). The average beta values for each cluster were then entered into a 3 × 4 repeated measures ANOVA. To test whether there was a significant difference between valence-specific activations across stimulus types, we examined the significance of the interaction and the main effect. Then we performed the follow-up tests for the significant interaction effect to test all stimulus types against each other within a specific valence by 3 separate ANOVAs, and also to test all valences against each other within a specific modality by 4 separate ANOVAs.

3. Results

3.1. Behavior analysis

Participants can recognize the valence successfully with high accuracy (mean accuracy = 93.8%, SD = 13.3%), indicating the valences can be recognized well regardless of different stimulus types. The accuracy and response time (RT) under the 12 conditions (3 valences × 4 stimulus types) were shown in Table 1. Considering participants can recognize the valence of almost all stimuli, we only conducted repeated measures ANOVA for the response time with the factors valences (positive, negative, and neutral) and stimulus types (face, body, voice, and music). The repeated measures ANOVA showed a significant interaction between valences and stimulus types ($F = 2.25$, $p = 0.04$, $\eta_p^2 = 0.06$), indicating that the RT varies with modalities at the different valences. The main effect of valence is not significant ($F = 1.75$, $p = 0.18$, $\eta_p^2 = 0.05$), while the effect of stimulus type is significant different ($F = 17.69$, $p < 0.001$, $\eta_p^2 = 0.20$). In addition, the one-way ANOVAs with emotional valences as factor suggested that the RTs were significantly different across different valences when the modality is the face expression ($F = 3.80$, $p = 0.03$), while there is no significant difference in the stimuli type of body ($F = 0.39$, $p = 0.68$), voice ($F = 0.90$, $p = 0.41$), and music ($F = 2.35$, $p = 0.10$).

Table 1
| The behavior accuracies and response times.

Valence	Modality	Accuracy (%)		Response time (ms)	
		Mean	SD	Mean	SD
Positive	face	99.23	2.72	757.00	180.11
	body	90.73	17.42	744.66	206.50
	voice	88.68	19.33	847.96	157.22
	music	98.85	4.31	757.73	197.80
Negative	face	94.79	11.99	809.49	185.67
	body	95.30	7.80	780.31	178.02
	voice	87.86	17.51	862.23	180.49
	music	86.50	18.95	848.17	185.94
Neutral	face	98.46	4.64	680.37	140.03
	body	96.92	7.36	739.92	153.88
	voice	98.08	6.34	805.10	138.86
	music	90.64	16.17	756.19	135.06

3.2. Searchlight analysis

In the searchlight analysis, we computed the correlation matrix for each voxel in the whole brain across all subjects. For the between-modality conditions, the correlation coefficients of within-valence and between-valence were averaged separately, and then the average coefficient of between-valence was subtracted from the average coefficient of within-valence. This procedure was repeated on each voxel, generating a difference map (d-map) for each subject. Then the group analysis was performed by one-sample *t*-test on all participants' d-maps to test whether the coefficient was non-zero. At last, four clusters were revealed: bilateral postcentral (−45, −12, 5 and 45, -21, 39), left middle temporal gyrus (MTG) (−57, −54, 15) and right middle frontal gyrus (MFG) (27, −6, 63), as shown in Fig. 3. Table 2 describes the details of these clusters in the group-level analysis.

3.3. Validation analyses

3.3.1. Cross-modal classification

The four clusters identified by the searchlight analysis were analyzed using cross-modal classification to examine whether these clusters truly provide information of modality-general valence, as suggested by Etzel et al.'s "Confirmation Test" (Etzel et al., 2013). The accuracies of the successful cross-modal classification of the three valences were significantly higher than the chance level (0.33) in the left postcentral gyrus (t (25) = 5.11, $p < 0.001$) and MTG (t (25) = 4.39, $p < 0.001$) but not for the right postcentral gyrus (t (25) = 0.26, $p = 0.80$) and MFG (t (25) = 0.48, $p = 0.80$) after performing FDR correction for multiple comparisons. The results of the classification can be found in Table 3. In addition, the details classification results in each stimulus type can be found in Table 4. The classification results showed that all the classification accuracies were significant in the four stimuli types in the left postcentral gyrus. However, the classification was only successful in the face and body condition in the left MTG.

3.3.2. Univariate analysis

The most obvious evidence of the representation of modality-general valence is that the cross-modal classification and searchlight analysis converge to the same result. So the clusters that cannot classify PvsNvs0 were excluded for further consideration. Univariate analysis was

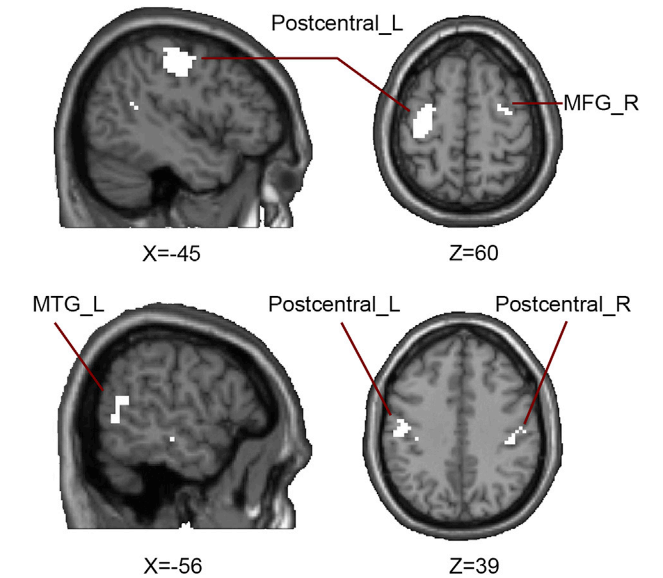

Fig. 3. Results of searchlight analysis showing the significant clusters which could represent valence across modalities. The searchlight analysis revealed modality-general representations of valence in four clusters: bilateral postcentral (−45, −12, 5 and 45, -21, 39), left middle temporal gyrus (MTG) (−57, −54, 15) and right middle frontal gyrus (MFG) (27, −6, 63), which are shown on sagittal and axial slices.

Table 2
| Significant clusters from searchlight analysis.

Anatomical region	Hemisphere	Cluster size	MNI coordinates			t-value
			x	y	z	
Postcentral	L	348	−45	−12	51	7.56
Postcentral	R	27	45	−21	39	4.63
MTG	L	33	−57	−54	15	4.73
MFG	R	21	27	−6	63	5.05

MTG: middle temporal gyrus; MFG: middle frontal gyrus; R, right; L, left.

Table 3
| Cross-modal classification accuracies for each searchlight cluster.

Clusters	Cross-modal classification			
	PvsNvs0			
	Accuracy (%)	SEM	$t_{(25)}$	$P_{(FDR\text{-}corrected)}$
Postcentral_L	38.46*	1.00	5.11	<0.001*
Postcentral_R	33.48	0.58	0.26	0.80
MTG_L	35.59*	0.51	4.39	<0.001*
MFG_R	33.56	0.51	0.45	0.80

One-sample t-tests of classification accuracies for PvsNvs0 in each cluster. *Significant P values. PvsNvs0, positive versus negative versus neutral.

performed to explore the differences in the activation under the 12 conditions. The beta values of different conditions were extracted from the left postcentral gyrus and MTG, and the detail results can be found in Fig. 4.

The repeated measures ANOVA was performed for the beta values with the factors valences and stimulus types. In the left postcentral gyrus, there are no significant interaction between these factors ($F = 0.83$, $p = 0.55$, $\eta_p{}^2 = 0.02$). In addition, the main effect of stimulus types is not significant ($F = 0.65$, $p = 0.59$, $\eta_p{}^2 = 0.01$), while the main effect of valences is significant ($F = 14.69$, $p < 0.001$, $\eta_p{}^2 = 0.28$). In addition, we performed pair comparisions between the valence categories averaged across stimulus types and the results suggested that the activation at different valence is significantly different [negative-positive ($t = 5.11$, $p < 0.001$), neutral-positive ($t = 6.62$, $p < 0.001$), neutal-negative ($t = 2.23$, $p = 0.035$)], showing a gradient that goes from positive to neutral passing through negative.

Nevertheless, the repeated measures ANOVA showed a significant interaction between valences and stimulus types ($F = 8.93$, $p < 0.001$, $\eta_p{}^2 = 0.19$) in left MTG, which indicated that the activation of three valences

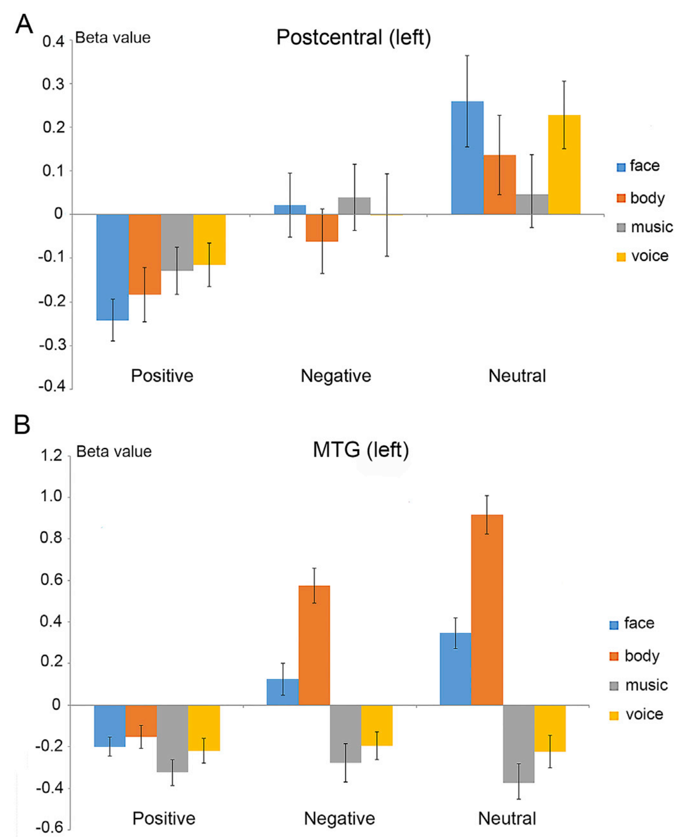

Fig. 4. Mean activation in left postcentral gyrus and left MTG. The mean beta values were used to showing the activation intensity of 12 conditions in the univariate analysis. Error bars show SEM.

Table 4
| Cross-modal classification accuracies in each stimuli type.

Stimuli type	Cross-modal classification							
	Postcentral_L				MTG_L			
	Accuracy (%)	SEM	$t_{(25)}$	P	Accuracy (%)	SEM	$t_{(25)}$	P
face	40.08*	1.24	5.43	<0.001*	37.02*	0.76	4.86	<0.001*
body	41.71*	1.19	7.05	<0.001*	39.20*	0.81	7.29	<0.001*
voice	36.90*	1.15	3.11	0.005*	33.50	0.79	0.21	0.84
music	35.17*	0.82	2.24	0.035*	32.85	0.69	−0.70	0.49

One-sample t-tests of classification accuracies for the left postcentral gyrus and MTG in each stimuli type. *Significant P values.

differed for four stimulus types. The main effect of stimulus types is significant ($F = 53.35$, $p < 0.001$, $\eta_p^2 = 0.41$). Then we performed the one-way ANOVA with stimulus types as factor within each specific valence, which suggested that the activation at different stimulus types is significant different when the valence is negative ($F = 24.33$, $p < 0.001$) and neutral ($F = 48.33$, $p < 0.001$). However, the difference is not significant in the positive valence ($F = 1.34$, $p = 0.27$). There is also significant difference between valences ($F = 21.30$, $p < 0.001$, $\eta_p^2 = 0.36$). Then one-way ANOVAs with valence categories as factor were conducted and the results suggested that the activation at different valence is significant different when the modality is the face expression ($F = 15.32$, $p < 0.001$) or body emotion ($F = 48.11$, $p < 0.001$). However, the difference is not significant in the voice ($F = 0.02$, $p = 0.98$) and music ($F = 0.66$, $p = 0.52$).

4. Discussion

The current study tried to identify brain regions representing emotion valence regardless of different stimulus types using three converging analysis techniques. First, applying the between-modality valence difference measure to the searchlight analysis at the whole-brain level, the bilateral postcentral gyrus, right MFG and left MTG were located. We subtracted the correlations of voxel activation patterns when modality and valence are different from the correlations when the modalities are different but the valence is the same. Therefore, this approach isolates the similarity of voxel activation patterns attributable to valence similarities. These regions may represent modality-general valence regardless of stimulus types in visual and auditory modalities Second, the cross-modal classification was utilized as the additional validation technique to the searchlight analysis. The results showed that the valence of PvsNvs0 could be categorized successfully across all the stimuli types in the left postcentral gyrus. While the valence of PvsNvs0 could be categorized only in face and body conditions in the left MTG. The classification results are unlikely to be driven by lower-level features since the lower-level features (e.g., brightness, motion, etc.) cannot be shared between different modalities. So the effects of lower-level features from one modality might not play a role in the testing of the other modality. At last, we adopted the univariate analysis to study the activation in the cluster derived from the cross-modal classification analysis. The activation of different stimulus types from the same valence is not significantly different in the left postcentral gyrus, but it is significantly different in the left MTG when the valence is negative and neutral. In addition, the activation of the same stimulus type showed a gradient that goes from positive to neutral passing through negative in the left postcentral gyrus (when the stimuli type is face, body, or voice) and MTG (when the stimuli type is face or body).

4.1. Representations of modality-general valence

The left postcentral gyrus was shown to be involved in emotional processing (Banziger et al., 2012; Flaisch et al., 2015; Kassam et al., 2013) and the perception of emotional face, body and vocal expressions (van de Riet et al., 2009). A recent study showed that the emotional categories were successfully decoded from different perceptual modalities (face and vocal expression) in the left postcentral gyrus (Kragel and LaBar, 2016). The area was also involved in the cross-modal representation when classifying four basic emotions caused by imagery through affective words and video clips (Saarimaki et al., 2016). These studies showed that the emotions across different modalities or types may share similar neurological characteristics, and the postcentral gyrus can represent emotions across different stimuli categories. Our previous work also suggested that the basic emotions can be represented regardless of different types of stimuli in the visual modality (face, body, and whole-person) in the left postcentral gyrus (Cao et al., 2018). Recent studies found that the postcentral gyrus was crucial for valence decoding. The activation pattern about valence caused by observing photos with different emotional categories was represented in the postcentral gyrus (Baucom et al., 2012). The work of Kim et al. also provided evidence of modality-general representations of the valences (negative, neutral and positive) in the left postcentral gyrus (Kim et al., 2017), which was consistent with our research. Video and music stimuli were used in this research, while our study used different stimulus types in visual and auditory modalities. The use of a broader range of stimuli categories is more in line with the channels through which we perceive emotions in our daily lives, which can contribute to generalizing the research about modality-general valence representations. Consistent with these previous studies, our current study provided further evidence for the role of the left postcentral gyrus in representing the modality-general valence by coding emotional information regardless of different stimulus types. And there is no significant difference in the activation of different stimulus types from the same valence in the left postcentral gyrus. In summary, the current study provided support for the representations of modality-general valence in the left postcentral gyrus and it suggested that the cluster could code valence information regardless of different types of stimuli in visual and auditory modalities.

4.2. Regions where valence could not be represented across different modalities

The MTG has been shown to accurately classify the basic emotions induced by movies and mental imagery, and it can also classify emotions across both modalities (Saarimaki et al., 2016). Kim et al. conducted a searchlight analysis of audio-visual emotional stimuli with different valence and found the MTG was located for representing valence, and the MTG was also shown to be critical for valence decoding (Kim et al., 2016). In the current study, however, the cross-modal classification was successful only in the stimuli types of face and body and the univariate results also showed MTG is only sensitive to valence in the visual domain. Therefore, the validation results suggested that the activation pattern from the left MTG may not be able to represent the modality-general valence but it can be involved in the valence representing across stimuli types in the visual modality.

The searchlight analysis also revealed that the right postcentral gyrus and MFG could be involved in representing modality-general valence regardless of stimulus types (face, body, voice or music). However, the classification of the emotional valences (PvsNvs0) failed based on the activation pattern in the two clusters when the cross-modal classification was utilized as the additional validation technique to the searchlight analysis. The results possibly arose from the overpowered nature of the searchlight technique, which might erroneously infer the existence of modality-general voxels that are not existent (Kim et al., 2017). And it may lead us to conclude that the hypothetical effect exists, but in fact, it does not exist. Based on the validation by the cross-modal classification, therefore, our findings suggested that the two brain regions may not carry enough information to represent the modality-general valence across the stimulus types.

The right postcentral gyrus and the MFG have been shown to be involved in the emotional processing of visual or auditory modalities (Sarkheil et al., 2013). The right postcentral gyrus was involved in the processing of emotional faces and the interaction effect between the type and intensity of emotions while decoding dynamic affective faces (Sarkheil et al., 2013). The right postcentral gyrus which corresponds to the primary somatosensory cortex has been shown to be critical for emotion recognition when decoding the emotional categories from facial and vocal cues (Adolphs, 2002). In the right postcentral gyrus, the emotional prediction pattern has been suggested to exhibit somatotopic organization, indicating that information related to the body state could help decode emotional expression (Kragel and LaBar, 2016). And this study further showed that facial mimicry or motoric participation may contribute to emotion recognition in the right postcentral gyrus, but are unlikely responsible for converging the perceived and subjective experiences, because of the observations that emotion prediction patterns in the primary motor cortex were unrelated to the behavioral self-report. However, most of the studies focused on the representations elicited from single face, body or voice or the representations of basic emotions. In addition, it also needs to be further studied whether the right postcentral gyrus is sensitive to the valence. Our research focused on the valence representation across different stimulus types, and the validation results suggested that the activation pattern from the right postcentral gyrus may not be able to represent the modality-general valence even though it can be involved in the emotional processing.

Similarly, the right MFG didn't show significant PvsNvs0 classification. The MFG has not been reported much in affective processing in previous studies. Specifically, positive and negative words cause increased activation in the MFG compared to neutral words (Cato et al., 2004; Herbert et al., 2009; Hoffmann et al., 2015; Kensinger and Schacter, 2009). Subramaniam et al. compared negative and neutral metaphors and found the increased signal on the right MFG (Subramaniam et al., 2013). A searchlight analysis was performed and revealed that the MFG was sensitive to valence (Kim et al., 2016). And Kim et al. have found a modality-general representation of valence in the MFG (Kim et al., 2017). Similar to our results, the MFG didn't show a significant PvsN classification in their study, but there are significant information from multidimensional scaling (MDS) results for PNvs0 and PvsN representations. We speculated that the broader range of stimuli categories in our experiment may be a factor causing the difference with the previous study. MDS method can be considered for the right postcentral gyrus and MFG to reveal the underlying mechanisms while the facial, bodily, vocal and musical expressions are perceived. Taken together, the searchlight analysis and validation analyses showed that although these brain regions may be involved in emotional processing, the right postcentral gyrus and the MFG might not represent the valence information across modalities where each modality contained different types of stimuli.

5. Conclusion

In our study, the searchlight analysis located four regions (the bilateral postcentral gyrus, left MTG and right MFG) that may be involved in the valence representation regardless of different stimulus types in the visual modality (face and body) and auditory modality (voice and music). By further validation, it revealed that only the left postcentral gyrus distinguish three valences across different stimulus types in different modalities, while left MTG is sensitive to the valence across stimulus types of the visual modality. In the future work, it can be considered for the right postcentral gyrus and MFG to reveal the underlying mechanisms while the facial, bodily, vocal and musical expressions are perceived.

Author contribution

BL and LC designed the experiment. LC performed the experiment. LC and JG analyzed results. LC wrote the manuscript. JG and BL contributed to the manuscript revision. All authors contributed to discuss the results and have approved the final manuscript.

Conflicts of interest

We declare that we have no actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations that can inappropriately influence our work.

Funding

This work was supported by the National Key Research and Development Program of China (No.2018YFB0204304), National Natural Science Foundation of China (No.U1736219 and No.61571327).

References

Adolphs, R., 2002. Neural systems for recognizing emotion. Curr. Opin. Neurobiol. 12, 169–177.

Banziger, T., Mortillaro, M., Scherer, K.R., 2012. Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. Emotion 12, 1161–1179.

Barrett, L.F., Russell, J.A., 2014. The Psychological Construction of Emotion. The Guilford Press.

Baucom, L.B., Wedell, D.H., Wang, J., Blitzer, D.N., Shinkareva, S.V., 2012. Decoding the neural representation of affective states. Neuroimage 59, 718–727.

Cao, L., Xu, J., Yang, X., Li, X., Liu, B., 2018. Abstract representations of emotions perceived from the face, body, and whole-person expressions in the left postcentral gyrus. Front. Hum. Neurosci. 12.

Cato, M.A., Crosson, B., Gökçay, D., Soltysik, D., Wierenga, C., Gopinath, K., Himes, N., Belanger, H., Bauer, R.M., Fischler, I.S., Gonzalez-Rothi, L., Briggs, R.W., 2004. Processing words with emotional connotation: an fMRI study of time course and laterality in rostral frontal and retrosplenial cortices. J. Cogn. Neurosci. 16, 167–177.

Chang, Chih-Chung, Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 (27), 21–27, 27.

Chikazoe, J., Lee, D.H., Kriegeskorte, N., Anderson, A.K., 2014. Population coding of affect across stimuli, modalities and individuals. Nat. Neurosci. 17, 1114–1122.

Eerola, T., Vuoskoski, J.K., 2010. A comparison of the discrete and dimensional models of emotion in music. Psychol. Music 39, 18–49.

Etzel, J.A., Zacks, J.M., Braver, T.S., 2013. Searchlight analysis: promise, pitfalls, and potential. Neuroimage 78, 261–269.

Flaisch, T., Imhof, M., Schmalzle, R., Wentz, K.U., Ibach, B., Schupp, H.T., 2015. Implicit and explicit attention to pictures and words: an fMRI-study of concurrent emotional stimulus processing. Front. Psychol. 6, 1861.

Fontaine, J.R., Scherer, K.R., Roesch, E.B., Ellsworth, P.C., 2007. The world of emotions is not two-dimensional. Psychol. Sci. 18, 1050–1057.

Herbert, C., Ethofer, T., Anders, S., Junghofer, M., Wildgruber, D., Grodd, W., Kissler, J., 2009. Amygdala activation during reading of emotional adjectives–an advantage for pleasant content. Soc. Cogn. Affect. Neurosci. 4, 35–49.

Hoffmann, M., Mothes-Lasch, M., Miltner, W.H., Straube, T., 2015. Brain activation to briefly presented emotional words: effects of stimulus awareness. Hum. Brain Mapp. 36, 655–665.

Kassam, K.S., Markey, A.R., Cherkassky, V.L., Loewenstein, G., Just, M.A., 2013. Identifying emotions on the basis of neural activation. PLoS One 8, e66032.

Kensinger, E., Schacter, D., 2009. Processing emotional picturesand words: effects of valence and arousal. Cogn. Affect. Behav. Neurosci. 6, 110–126.

Kim, J., Schultz, J., Rohe, T., Wallraven, C., Lee, S.W., Bulthoff, H.H., 2015. Abstract representations of associated emotions in the human brain. J. Neurosci. 35, 5655–5663.

Kim, J., Shinkareva, S.V., Wedell, D.H., 2017. Representations of modality-general valence for videos and music derived from fMRI data. Neuroimage 148, 42–54.

Kim, J., Wang, J., Wedell, D.H., Shinkareva, S.V., 2016. Identifying core affect in individuals from fMRI responses to dynamic naturalistic audiovisual stimuli. PLoS One 11, e0161589.

Klasen, M., Kenworthy, C.A., Mathiak, K.A., Kircher, T.T., Mathiak, K., 2011. Supramodal representation of emotions. J. Neurosci. 31, 13635–13643.

Koelsch, S., Fritz, T., DY, V.C., Muller, K., Friederici, A.D., 2006. Investigating emotion with music: an fMRI study. Hum. Brain Mapp. 27, 239–250.

Kragel, P.A., LaBar, K.S., 2015. Multivariate neural biomarkers of emotional states are categorically distinct. Soc. Cogn. Affect. Neurosci. 10, 1437–1448.

Kragel, P.A., LaBar, K.S., 2016. Somatosensory representations link the perception of emotional expressions and sensory experience. eNeuro 3.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A 103, 3863–3868.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540.

Liang, Y., Liu, B., Li, X., Wang, P., 2018. Multivariate pattern classification of facial expressions based on large-scale functional connectivity. Front. Hum. Neurosci. 12.

Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal representations of perceived emotions in the human brain. J. Neurosci. 30, 10127–10134.

Saarimaki, H., Gotsopoulos, A., Jaaskelainen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., Nummenmaa, L., 2016. Discrete neural signatures of basic emotions. Cerebr. Cortex 26, 2563–2573.

Sarkheil, P., Goebel, R., Schneider, F., Mathiak, K., 2013. Emotion unfolded by motion: a role for parietal lobe in decoding dynamic facial expressions. Soc. Cogn. Affect. Neurosci. 8, 950–957.

Shinkareva, S.V., Wang, J., Kim, J., Facciani, M.J., Baucom, L.B., Wedell, D.H., 2014. Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data. Hum. Brain Mapp. 35, 3558–3568.

Subramaniam, K., Beeman, M., Faust, M., Mashal, N., 2013. Positively valenced stimuli facilitate creative novel metaphoric processes by enhancing medial prefrontal cortical activation. Front. Psychol. 4, 211.

van de Riet, W.A., Grezes, J., de Gelder, B., 2009. Specific and common brain regions involved in the perception of faces and bodies and the representation of their emotional expressions. Soc. Neurosci. 4, 101–120.

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., Belin, P., 2014. Crossmodal adaptation in right posterior superior temporal sulcus during face-voice emotional integration. J. Neurosci. 34, 6813–6821.

Woo, C.W., Krishnan, A., Wager, T.D., 2014. Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. Neuroimage 91, 412–419.

Zhang, G., Cheng, Y., Liu, B., 2016. Abnormalities of voxel-based whole-brain functional connectivity patterns predict the progression of hepatic encephalopathy. Brain. Image. Behave. 11, 784–796.