Neural Activity of Different Emotion Modalities and Their Similarities and Differences: An ALE Meta-Analysis of Brain Imaging Studies

Authors: Liu Juncai, Ran Guangming, Zhang Qi, Ran Guangming

Date: 2021-10-12T00:00:00+00:00

Abstract

Emotion recognition has long been a focus of scholarly attention. Although previous studies have investigated the neural mechanisms of dynamic facial expressions, dynamic body expressions, and vocal emotions, the overall understanding of each emotion modality remains relatively incomplete, and the commonalities and distinctions in neural mechanisms across different emotion modalities are poorly understood. Therefore, this study first identified brain activation regions for each emotion modality through three independent activation likelihood estimation meta-analyses, and then conducted comparative analyses to evaluate common and unique neural activities across the three emotion modalities. The results showed that brain activity for dynamic facial expressions includes extensive frontal, occipital, temporal, and partial parietal cortices, as well as subcortical regions such as the hippocampus, cerebellum, thalamus, and amygdala; activation for dynamic body expressions was concentrated in temporal/occipital-related brain regions, as well as the cerebellum and hippocampus; vocal emotions elicited activation in the temporal lobe, frontal lobe, amygdala, caudate nucleus, and insula. Conjunction analysis revealed that the three emotion modalities cross-modally activated the left middle temporal gyrus and right superior temporal gyrus. Comparative analysis results demonstrated the predominance of visual over auditory stimuli, with dynamic facial expressions being particularly prominent, while dynamic body expressions also played a significant role, though vocal emotions exhibited distinct characteristics. Taken together, these findings validate and extend existing neural models of the three emotion modalities, revealing central, universal regions for emotion processing, yet each emotion modality possesses its own reliable and specific neural circuits.

Neural Activity of Different Emotion Carriers and Their Similarities and Differences: An ALE Meta-Analysis of Neuroimaging Studies

Liu Juncai[1], **Ran Guangming**[1], **Zhang Qi**[2]
(1. Department of Psychology, School of Education, China West Normal University, Nanchong 637002;
2. College of Preschool and Primary Education, China West Normal University, Nanchong 637002)

Abstract

Emotion recognition has long been a focal topic in psychological research. While previous studies have explored the neural mechanisms underlying dynamic facial expressions, dynamic bodily expressions, and emotional voices, our comprehensive understanding of each emotion carrier remains relatively incomplete, and the commonalities and distinctions among their neural mechanisms are poorly understood. Therefore, this study first employed three independent activation likelihood estimation (ALE) meta-analyses to identify brain regions activated by each emotion modality, followed by comparative analyses to evaluate shared and unique neural activity across the three emotion carriers. The results revealed that dynamic facial expressions engaged extensive frontal, occipital, temporal, and partial parietal cortices, as well as subcortical regions including the hippocampus, cerebellum, thalamus, and amygdala. Dynamic bodily expressions showed activation concentrated in temporal/occipital regions, cerebellum, and hippocampus. Emotional voices elicited activation in temporal and frontal lobes, amygdala, caudate nucleus, and insula.

Conjunction analysis indicated that the three emotion carriers cross-modally activated the left middle temporal gyrus and right superior temporal gyrus. Contrast analysis demonstrated that visual stimuli held an advantage over auditory stimuli, with dynamic facial expressions being particularly prominent, though dynamic bodily expressions also played an important role, while emotional voices exhibited distinct characteristics. Overall, these findings validate and extend existing neural models of the three emotion carriers, revealing central, universal regions for emotion processing while demonstrating that each emotion carrier relies on its own reliable, specific neural circuits.

Keywords: emotion carriers, dynamic facial expressions, dynamic bodily expressions, emotional voices, ALE meta-analysis

As a highly social species, humans constantly engage in frequent social interactions. Accurately perceiving others' emotions represents one of the most crucial social skills in these interactions. The psychological community has long focused on facial expressions, body movements, and vocal emotions because they constitute indispensable communicative signals in establishing and maintaining social relationships. For instance, fear can be recognized through visual information from the eye region of faces, body postures and movements, and acoustic changes in voices (de Gelder et al., 2006; Heberlein & Atkinson, 2009). Despite these signals being fundamentally different, they can all lead to identical emotion recognition (i.e., that a person is afraid) and activate similar emotion-specific responses in observers (Magnée et al., 2007).

Received: October 6, 2021
*Supported by the National Natural Science Foundation Youth Project (31900760) and the Provincial College Student Innovation and Entrepreneurship Training Program Project (S201910638117).

Corresponding author: Ran Guangming, E-mail: haiqi198649@163.com

Dynamic Facial Expressions and Their Neural Basis

The face represents the most common stimulus form in social life, with over 100 anatomically and functionally distinct muscles creating an enormous space of possible configurations (Kilts et al., 2003; Waters & Terzopoulos, 1990). Subtle muscular changes provide rich and powerful social information, allowing us to rapidly and accurately infer others' traits from facial features (Ritchie et al., 2017) and automatically form impressions of others (Klapper et al., 2016) to facilitate further social communication (Haxby et al., 2000; Zhang et al., 2015). In daily life, however, we encounter dynamically changing rather than static faces, making accurate recognition of others' dynamic emotional faces even more critical for individual survival and interpersonal interaction. Research indicates that a core neural network exists for processing dynamic facial expressions, namely the temporo-occipital junction (Foley et al., 2012; Haxby et al., 2000), which includes the fusiform gyrus (FG), superior temporal gyrus (STG), middle temporal gyrus (MTG), inferior temporal gyrus (ITG), middle occipital gyrus (MOG), and inferior occipital gyrus (IOG). The extended network beyond this core includes widespread cortical and subcortical regions: cortical areas such as the inferior parietal lobule (IPL) (Sarkheil et al., 2013) and inferior frontal gyrus (IFG) (Sato et al., 2012), and subcortical regions primarily involving the amygdala (Sato et al., 2010). The core network is mainly responsible for early visual analysis and motion processing of stimuli, while the extended network is associated with affective processing, facial mimicry, and dynamic representation of stimuli (Haxby et al., 2000; Zhang et al., 2015). Zinchenko et al.' s (2018) meta-analysis found that dynamic facial expressions activated the aforementioned brain regions plus the cerebellum, validating and supporting the neural network for dynamic facial expression processing.

Dynamic Bodily Expressions and Their Neural Basis

Beyond facial expressions, bodily expressions constitute another important carrier of emotional communication. For a long time, facial expressions were considered more universal and consistent carriers of emotional information than bodily expressions (de Gelder et al., 2014). However, research shows that adults can recognize emotions from body postures and movements with accuracy comparable to that from faces (Atkinson et al., 2007; Zieber et al., 2014), and in some cases, the body can convey emotional states even better than the face (de Gelder & Partan, 2009; Sinke et al., 2012). Compared to faces, the body offers unique advantages for emotional communication: its relatively larger size makes it visible from greater distances, and emotional signals can be observed even from behind the signaler (de Gelder, 2009). Previous research on the processing mechanisms of dynamic bodily expressions has held differing views. Some researchers endorse Giese and Poggio' s (2003) hierarchical neural model for body form and motion processing, proposing that the processing pathway for body form information (e.g., shape features) begins in visual cortex, passes through secondary temporal cortex, and reaches the fusiform gyrus and superior temporal sulcus (STS), while the pathway for motion and emotion information originates in visual cortex, passes through occipital and temporal lobes, and ultimately reaches the STS, which integrates complex human action information (de Gelder et al., 2014; Moro et al., 2008). However, most researchers favor de Gelder et al.' s (2006, 2015) dual-route theory for bodily expression processing. This theory posits two independent yet cooperative neural pathways: an automatic processing pathway for bodily expressions that supports reflexive behavior, primarily involving the superior colliculus (SC), thalamus, striatum, and amygdala; and a cortical control circuit responsible for affective, cognitive, and decision-making processes, including the frontoparietal motor system and prefrontal regions connected to the amygdala.

Vocal Emotions and Their Neural Basis

Voices also carry diverse information about the actor (Krauss et al., 2002; Pisanski et al., 2016) and are as easily recognizable as faces and bodies (Scherer, 1995). Unlike visual stimuli with higher spatial sensitivity, voices primarily enable us to discriminate finer temporal differences (Calvert et al., 1998). Compared to faces and bodies, voices are more hierarchical (diverse), with intensity, speed, F0, and other features all conveying information (Scherer et al., 2003). Researchers propose that voices mainly transmit speech, emotion, and identity information via three partially separable neural pathways (Belin et al., 2004). The neural mechanism for speech information processing can be well understood through the dual-pathway model (Hickok & Poeppel, 2007). In this model, the middle and posterior superior temporal sulcus (M/PSTS) in both hemispheres processes phonological representation of vocal stimuli, then splits into ventral and dorsal pathways for semantic comprehension and conversion of phonological representations encoded in the STS into articulatory motor representations encoded in motor regions (IFG), respectively. Emotion information processing can be well explained by a three-stage working model (Schirmer & Kotz, 2006): around 100 ms post-stimulus, primary and secondary auditory cortices extract and analyze acoustic information; around 200 ms, the anterior superior temporal gyrus/sulcus (ASTG/S) and amygdala integrate emotional cues; around 400 ms, the right IFG and orbitofrontal cortex (OFC) perform higher-level cognitive evaluation, while the left inferior frontal cortex processes semantic emotional information. Identity information processing can be comprehensively understood through an integrative model (Maguinness et al., 2018). According to this model, processing at the perceptual level involves analyzing identity features, primarily engaging posterior STS, Heschl' s gyrus (HG), and planum temporale (PT); at the recognition level, extracted identity features are compared with stored voice prototypes in the middle STS, selecting features that deviate from prototypes; subsequently, anterior and middle STS compare deviant features with stored reference schemas for voice-identity recognition. Meta-analytic studies have validated these theories while extending the roles of the cerebellum and insula (Frühholz & Grandjean, 2013; Frühholz et al., 2016).

Similarities and Differences in Neural Mechanisms Across Emotion Carriers

Although numerous studies have examined the neural mechanisms of the three emotion carriers, most have focused on unimodal processing. With deepening understanding, some researchers have begun investigating similarities and differences between facial and bodily expression processing mechanisms. On one hand, studies have found overlapping neural mechanisms between facial and bodily expression processing. Both static and dynamic bodily expressions can significantly activate face-sensitive regions such as the fusiform gyrus, STS, and amygdala (Kret et al., 2011; Giese & Poggio, 2003). Additionally, facial and bodily stimuli of different emotion types can co-activate emotion-processing regions including the OFC (de Gelder, 2006), insula (Tamietto et al., 2015), and cuneus (Kret et al., 2011). On the other hand, evidence supports dissociable neural mechanisms for face and body expression processing. Downing et al. (2001) directly compared different stimuli and found that the extrastriate body area (EBA) was sensitive only to bodily stimuli, not to faces or other stimuli. Moreover, although peak activation locations overlapped between facial and bodily expressions, specific activation sites were dissociated (Schwarzlose et al., 2005; van de Riet et al., 2009).

Beyond examining similarities and differences between facial and bodily expressions, researchers have also identified commonalities and distinctions between facial expressions and vocal emotions. For example, Dricu and Frühholz' s (2016) meta-analysis showed that both faces and voices activated bilateral inferior frontal cortex (IFC), dorsal medial frontal cortex (DMFC), right STS, and left insula; faces showed greater activity than voices in bilateral amygdala, middle fusiform gyrus (MFG), and visual association areas; whereas voices elicited greater activation than faces in right middle STS, left IFC, left medial frontal cortex, and frontal operculum (FOP). Schirmer' s (2018) meta-analysis matched included facial and vocal studies: for each vocal study, a task- and emotion-matched facial study was selected; vocal studies without modality matches were excluded; if one vocal study had multiple matching facial studies, further matching was performed based on sample size, gender ratio, and publication date, selecting the study with minimal differences. The results revealed that faces showed greater activity than voices in bilateral parahippocampal gyrus (PHG), left subcallosal gyrus (SG), and right putamen; voices showed greater activation than faces in bilateral STS.

Although most previous investigations of similarities and differences across emotion carriers have focused on two carriers, 极少数 researchers have examined similarities and differences among multimodal stimuli. Peelen et al. (2010) asked participants to evaluate emotional intensity perceived from facial movements, body actions, and vocal intonation, finding that the medial prefrontal cortex (MPFC) and left posterior STS could represent basic emotions independent of emotional stimulus carrier. Cao Linjing (2018) conducted different experiments: Experiment 1 required participants to judge emotion types of visual stimuli regardless of stimulus category (face, body, full-body emotion), revealing that only the left postcentral gyrus (POCG) could significantly discriminate three emotions cross-modally; Experiment 2 required valence judgments (negative, neutral, positive) for four stimulus categories (face, body, voice, music) across visual and auditory modalities, finding that left POCG and MTG could represent valence information unrestricted by both modality (visual and auditory) and stimulus category within each modality (face/body for visual, voice/music for auditory).

Limitations of Previous Research

On one hand, empirical studies have inherent unavoidable limitations. First, individual neuroimaging studies typically use relatively small sample sizes, potentially resulting in low statistical power and effect sizes (Yarkoni, 2009). Second, neuroimaging data are particularly sensitive to choices of tasks and control conditions, potentially leading to inconsistent results (Zhou et al., 2017). Therefore, it is necessary to employ meta-analytic techniques based on large-scale data synthesis to overcome limitations of individual neuroimaging studies (Eickhoff et al., 2009; Yarkoni et al., 2011). This not only helps compensate for incomplete understanding of each emotion carrier (for example, beyond the aforementioned brain regions, numerous studies have found that dynamic facial expressions and vocal emotions activate additional brain regions (Brattico et al., 2011; Sato et al., 2012; Trautmann et al., 2009). Moreover, processing models for bodily expressions remain debated without unified conclusions), but also enables investigation of generalities and differences in neural activity across the three emotion carriers, and provides representative reference coordinates for future region-of-interest (ROI) analyses. On the other hand, existing meta-analyses have certain limitations. First, in meta-analyses of dynamic facial expressions, most data came from female participants, and included studies investigated both rigid facial motion (head rotation relative to the body, such as nodding/shaking) and non-rigid motion (movement of facial features, such as blinking/pouting), potentially introducing bias. Second, no meta-analysis has yet been conducted on dynamic bodily expressions to identify their brain activation patterns. Third, meta-analyses of vocal emotions have included non-human environmental sounds, and some studies have focused only on emotional prosody processing while neglecting musical emotion and intonation. Additionally, although some meta-analyses have compared faces and voices, they did not distinguish between static and dynamic faces, limiting result reliability and generalizability. Finally, no meta-analytic studies have compared faces vs. bodies or bodies vs. voices, leaving similarities and differences among the three emotion carriers' neural mechanisms relatively unclear.

Purpose of the Current Meta-Analysis

In summary, current understanding of each emotion carrier remains relatively incomplete, and commonalities and distinctions among different emotion carriers' brain mechanisms are poorly understood. Therefore, this study first employed activation likelihood estimation (ALE) meta-analysis to identify brain regions activated by each emotion modality. Then, comparative analyses of these meta-analytic activation maps were conducted to evaluate shared and unique neural mechanisms across the three emotion carriers.

Literature Search

Studies on the brain mechanisms of emotion published in PubMed, PsycINFO, CNKI, and Google Scholar were searched (deadline: October 10, 2020). Keywords were grouped into four categories: (1) neuroimaging-related: neuroimaging, functional magnetic resonance imaging (fMRI), positron emission tomography (PET), Neuroimaging, fMRI, PET; (2) stimulus carrier-related: face, body, voice, prosody, intonation, music, facial, body, voice, vocal, prosody, intonation, music; (3) motion state-related: dynamic, move, motion, dynamic, move, motion. For faces and bodies, all three keyword groups were paired, while for vocal emotions, only the first two groups were paired. Additionally, forward and backward citation searches were performed on all identified studies to ensure comprehensive inclusion. This search strategy yielded 8,494 articles.

Literature Selection Criteria

Inclusion and exclusion criteria were as follows: (1) Study type: Empirical studies only; reviews, meta-analyses, and case studies were excluded. (2) Participants: Healthy individuals (mean age 9–55 years); patients with brain damage or psychological disorders and special populations, as well as children under 9 or adults over 55, were excluded. (3) Stimuli: Human dynamic faces (non-rigid motion), dynamic bodies, and voices expressing one of six basic emotions or neutral states; studies using static stimuli, rigid facial motion, point-light bodies, animal sounds, or non-basic emotions/neutral states were excluded. (4) Methodology: fMRI or PET only; studies using cognitive behavioral, EEG, MEG, or DTI/MRI analyzing white/gray matter were excluded. (5) Data analysis: Whole-brain analysis only; ROI-only studies were excluded. (6) Results reporting: Studies reporting activation coordinates in standardized Talairach or MNI space; studies without reported coordinates were excluded. The specific screening process and results are shown in Figure 1 [Figure 1: see original paper].

Literature Coding

The following information was coded from included studies: publication details (first author and year), sample size (total and male participants), age (mean or range), coordinate space (Talairach or MNI), imaging method (fMRI or PET), specific experimental contrasts, and coordinate values. The first author coded the literature twice according to inclusion/exclusion criteria, and the corresponding author verified the two coding results. The results showed high consistency with only minor data deviations. Basic information for included studies is shown in Table 1 .

Activation Likelihood Estimation

This study employed the most commonly used meta-analytic method in neuroimaging—Activation Likelihood Estimation. ALE uses Gaussian probability density distributions to model localization uncertainty of activation foci; the voxelwise union of these distributions produces ALE values, representing the estimated likelihood that at least one focus in the dataset is located within a given voxel (Turkeltaub et al., 2012; Morrison, 2016). Data processing was conducted using GingerALE software (version 3.0.2) (Eickhoff et al., 2009; Turkeltaub et al., 2012). Before analysis, coordinates reported in MNI space were converted to Talairach space using the software' s conversion tool. Separate meta-analyses were then performed for the three emotion carriers. Statistical maps were thresholded at $p < 0.05$ using family-wise error correction (cluster-level FWE), with cluster-forming threshold at $p < 0.001$ and 5,000 permutations for multiple comparisons correction (Eickhoff et al., 2017). Finally, pairwise contrast analyses were conducted across the three emotion carriers. Consistent with previous research (Arsalidou et al., 2020; Yalpe & Arsalidou, 2018), contrast analysis thresholds were set at $p < 0.01$ with minimum cluster size > 200 mm^3 (10,000 permutations).

Results

Compared to static carriers, dynamic faces showed greater activation in bilateral STG, MTG, FG, left ITG, PHG, amygdala, right IFG, precentral gyrus (PG), and lentiform nucleus (LN). Dynamic bodies were more active in bilateral STG and right cerebellum. Relative to neutral stimuli, dynamic emotional faces activated bilateral STG, MTG, FG, cerebellum, and amygdala, left ITG and MOG, right IFG, medial frontal gyrus (MFG), PG, insula, and IPL. Dynamic emotional bodies activated bilateral STG, left MOG, IOG, cuneus, cerebellum, right ITG, FG, and PHG. Emotional voices showed greater activation in bilateral STG, left MTG, transverse temporal gyrus (TTG), PG, and amygdala, and right IFG, MFG, and insula. Compared to neutral and negative carriers, dynamic positive faces were more active in right MTG, IOG, and cerebellum; dynamic positive bodies activated left cerebellum and right STG, MOG, and lingual gyrus (LING); positive vocal emotions activated bilateral STG, left amygdala, right TTG, and caudate nucleus (CN). Negative emotion effects revealed that dynamic negative faces activated bilateral MTG and amygdala, left IFG, right STG, superior frontal gyrus (SFG), MOG, FG, PG, and PHG; dynamic negative bodies were more active in left STG, MOG, IOG, and cuneus, and right MTG and FG; negative voices activated bilateral STG and IFG, left TTG, PG, and right insula.

Conjunction Analysis Results

Conjunction analysis results are presented in Table 2 and Figure 2 [Figure 2: see original paper]. Activation maps in columns 1–6 of Figure 2 show brain regions commonly activated between pairs of emotion carriers. Under dynamic effects, dynamic faces and bodies co-activated bilateral STG, left FG, and right cerebellum. Under emotion effects, dynamic faces and bodies co-activated bilateral STG, left MTG, IOG, and right ITG, FG, PHG, and LN; dynamic faces and voices co-activated left MTG, amygdala, and right STG, IFG, and insula; dynamic bodies and voices co-activated left MTG and right STG. Under positive emotion effects, dynamic faces and bodies co-activated in right ITG and MTG. Under negative emotion effects, dynamic faces and bodies co-activated bilateral ITG, left MTG, and right IOG. No common activation regions were found between dynamic faces and voices or between dynamic bodies and voices under positive or negative emotion effects. Column 7 shows all brain regions commonly activated between dynamic faces and bodies across dynamic, emotion, positive, and negative effects. Columns 8–11 show further overlays of commonly activated regions between pairs of emotion carriers, with overlapping regions representing common activation. Notably, the regions commonly activated across all three emotion carriers are located in left MTG and right STG.

Contrast Analysis Results

Contrast analysis revealed that under dynamic effects, bodies activated left STG more than faces. Under emotion effects, faces activated right FG, PG, PHG, and cerebellum more than bodies, while bodies activated left IOG, LING, cuneus, and right MOG more than faces. Compared to voices, faces activated bilateral PHG, left cerebellum and amygdala, and right STG, SFG, FG, PG, and IPL more strongly, while voices activated bilateral STG and left insula. Bodies activated left ITG, LING, cuneus, and right MTG, FG, and LN more than voices, whereas voices activated bilateral STG and right MTG, PG more than bodies. Under positive emotion effects, bodies activated right MOG more than faces, while faces activated right ITG, MOG, IOG, FG, and cerebellum more than bodies. Compared to faces, voices activated bilateral insula and left STG more, while bodies activated right MOG. Under negative emotion effects, bodies activated left cuneus, right MOG, and cerebellum more than faces. Compared to voices, faces activated bilateral MTG, left IFG and amygdala, and right FG and PG more strongly, while voices activated bilateral STG and left TTG more than faces. Bodies activated bilateral ITG and FG, and left STG, MTG, and cuneus more than voices, whereas voices activated bilateral STG, TTG, and left PG and insula more than bodies (see Table 3 and Figure 3 [Figure 3: see original paper]). No brain regions showed greater activation for dynamic faces than dynamic bodies under dynamic, positive, or negative emotion effects, and no regions showed greater activation for vocal emotions than dynamic bodies under negative emotion effects.

Discussion

This study used ALE meta-analysis to identify brain regions activated by dynamic facial expressions, dynamic bodily expressions, and emotional voices, and to elucidate common and distinct brain mechanisms across the three emotion carriers. In the following sections, we discuss the implications of these findings for understanding emotion recognition, focusing on similarities and differences across emotion carriers.

Dynamic Facial Expressions Consistent with previous meta-analyses of dynamic facial expressions (Zinchenko et al., 2018), this study found activation in bilateral FG and MTG, left MOG, cerebellum, amygdala, and IFG, and right STG and IOG. Additionally, bilateral PHG, left STG and ITG, and right MOG, SFG, MFG, IFG, PG, IPL, cerebellum, amygdala, LN, and insula were active. Bilateral STG, MTG, FG, left ITG, PHG, amygdala, and right IFG, PG, and LN were sensitive to dynamic information; bilateral STG, MTG, FG, cerebellum, and amygdala, left ITG and MOG, and right IFG, MFG, PG, insula, and IPL were more active for emotional information; right MTG, IOG, and cerebellum were more sensitive to positive emotional information; and bilateral MTG and amygdala, left IFG, and right STG, SFG, MOG, FG, PG, and PHG were more active for negative emotional information. These findings collectively validate and extend the core-extended neural network model for dynamic facial expression processing (Zhang et al., 2015). The results indicate that the core neural network (temporo-occipital junction) primarily includes FG, STG, MTG, ITG, MOG, and IOG. The extended network comprises two components: first, widespread cortical regions beyond the core network, including SFG, MFG, IFG, PG, and IPL; second, related subcortical regions involving PHG, cerebellum, amygdala, LN, and insula. The core network is mainly responsible for early perceptual encoding and motion processing of stimuli, while the extended network is associated with spatial attention, facial mimicry, emotion processing, and dynamic representation of stimuli. These results substantially refine the neural network for dynamic facial expression processing, providing a foundational model for dynamic facial expression recognition.

Dynamic Bodily Expressions As previously mentioned, recent research has validated and supported de Gelder et al.’ s (2006, 2015) dual-route model. However, our results—showing that dynamic bodily information activated bilateral STG and right cerebellum; emotion information activated bilateral STG, left MOG, IOG, cuneus, cerebellum, and right ITG, FG, and PHG; positive emotion information activated left cerebellum and right STG, MOG, and LING; and negative emotion information activated left STG, MOG, IOG, and cuneus and right MTG and FG—appear to better support Giese and Poggio' s (2003) hierarchical neural model for body form and motion processing. This is not contradictory, as Giese and Poggio' s theory emphasizes inferring emotional information from body form and motion through a bottom-up mechanism without involving higher cognitive processes such as emotion classification and modulation based on prior experience and memory. In contrast, de Gelder et al.' s (2006, 2015) theory combines bottom-up and top-down processing, emphasizing both automatic processing and cognitive modulation of emotion and behavior. The current meta-analysis did not investigate higher cognitive processing effects, thus no activation in related brain regions was found. This study presents a more comprehensive dynamic bodily expression processing model based on Giese and Poggio' s (2003) hierarchical neural model: the processing pathway for body form information begins in visual cortex, passes through secondary temporal cortex, and reaches the FG and STS; the pathway for motion and emotion information also originates in visual cortex, passes through extensive occipito-temporal cortex, and ultimately reaches the STS, which integrates complex human action information. Meanwhile, subcortical regions including the cerebellum, cuneus, LING, and PHG are also differentially involved in visual, motion, and emotion information processing.

Vocal Emotions This study found that emotional voices activated bilateral STG, right IFG, MFG, and insula, and left MTG, TTG, PG, and amygdala; positive vocal emotions activated bilateral STG, left amygdala, right TTG, and CN; negative vocal emotions activated bilateral STG and IFG, right insula, left TTG, and PG. These findings are largely consistent with previous meta-analyses on vocal emotions (Frühholz & Grandjean, 2013; Frühholz et al., 2016), supporting current understanding of the neural basis of vocal emotions: (1) Specific regions for vocal emotion processing exist in human auditory cortex, namely the temporal voice areas (TVA) distributed along bilateral STG/STS. (2) The brain processes speech, emotion, and identity information via three partially separable pathways: speech information is processed in STS, posterior lateral temporal cortex, anterior temporal cortex, and IFG; emotion information is processed in primary and secondary auditory cortices, STG/STS, amygdala, OFC, and inferior frontal cortex; identity information is represented in STG/STS, HG, PT, and MTG. The findings also extend understanding of MFG, insula, and CN involvement in vocal emotion processing (Bestelmeyer et al., 2014; Mitterschiffthaler et al., 2007; Numminen-Kontti, 2014). Additionally, we discovered activation in the PG not reported in previous meta-analyses. Prior research has found that PG, as part of frontal emotion-processing regions, is associated with both positive and negative emotion processing (Brattico et al., 2011; Fecteau et al., 2007; Numminen-Kontti, 2014).

Conjunction Analysis: Common Neural Mechanisms Across Emotion Carriers

Conjunction analysis revealed that dynamic faces and bodies co-activated bilateral STG, MTG, ITG, IOG, and FG, plus right PHG, LN, and cerebellum; dynamic faces and vocal emotions co-activated left MTG, amygdala, and right STG, IFG, and insula; dynamic bodies and vocal emotions co-activated left MTG and right STG. Notably, left MTG and right STG were commonly activated across all three emotion carriers.

Left Middle Temporal Gyrus The middle temporal gyrus is a crucial brain region with complex and diverse functions. Xu et al. (2019) used resting-state functional connectivity and co-activation patterns to characterize distinct MTG functions, finding that left MTG plays roles in social cognition, phonological processing, action observation, emotion perception, theory of mind, voice recognition, and face monitoring/recognition. Meta-analyses indicate that left MTG plays an important role in dynamic facial expression processing (Arsalidou et al., 2011; Zinchenko et al., 2018), primarily responsible for characteristic facial motion. Left MTG activation has also been found in dynamic bodily expression studies, involving action perception, selective activation for body shape, and emotion recognition (Grèzes et al., 2007; Jastorff et al., 2015; Poyo Solanas et al., 2020). Vocal emotion research has found that left MTG is not only selectively activated by voices (Ceravolo, 2013) and sensitive to speech information (Binder et al., 1997), but is also an important region for emotion processing, including emotion-neutral comparisons (Leitman et al., 2010; Regenbogen et al., 2012), emotion-emotion contrasts (Numminen-Kontti, 2014), and complex emotion understanding (Alba-Ferrara et al., 2011). Cao Linjing (2018) used dynamic faces, dynamic bodies, voices, and music stimuli and found that left MTG could represent emotional information cross-modally (visual and auditory) and 不受视觉（动态面孔和动态身体）和听觉（声音和音乐）模态内各自不同刺激类别的限制. In summary, left MTG plays a universal role in expression recognition, being activated in both unimodal and multimodal stimulus comparisons, though its specific functions differ across carriers.

Right Superior Temporal Gyrus Over the past two decades, extensive research on the STG and surrounding areas has paradoxically made its role in the human brain increasingly enigmatic. Bukowski and Lamm (2018) systematically summarized relatively well-established STG functions, including speech perception (Hickok & Poeppel, 2007), voice recognition (Kriegstein & Giraud, 2004), music perception (Mitterschiffthaler et al., 2007), face processing (Zhang et al., 2015), biological motion (Pelphrey et al., 2005), audiovisual integration (Stevenson & James, 2009), autobiographical memory and resting-state activity (Spreng & Grady, 2010), self-other attribution (Sperduti et al., 2011), theory of mind, empathy, and personality (Olson et al., 2007). Research has found that right STG is activated not only by single-modality stimuli but also co-activated by dynamic faces and bodies (Goldberg et al., 2015; Kret et al., 2011) and by dynamic faces and vocal emotions (Dricu & Frühholz, 2016; Watson et al., 2014). A limited portion of the right posterior STG shows integrative preference for human information compared to objects (Watson et al., 2014). Additionally, disease studies of schizophrenia, autism, and ADHD have shown that right posterior STG is associated with emotion recognition deficits (Alaerts et al., 2013; Mier et al., 2016; Zuberer et al., 2020). The current study further demonstrates the important role of right STG in expression recognition [understanding and recognizing emotions after sensory processing of stimuli (Dricu & Frühholz, 2016; Yang et al., 2015)], and indicates that right STG can represent emotions across dynamic faces, dynamic bodies, and vocal emotions.

Contrast Analysis: Divergent Neural Mechanisms Across Emotion Carriers

Although faces, bodies, and voices are all indispensable communicative signals in establishing and maintaining social relationships and can lead to identical emotion recognition, each emotion carrier has its own reliable specific activation regions. Overall, dynamic faces were particularly prominent, eliciting more extensive brain activity than dynamic bodies and vocal emotions, involving bilateral MTG, FG, PHG, cerebellum, left IFG and amygdala, and right STG, ITG, SFG, PG, and IPL. Compared to the other two carriers, dynamic bodies engaged bilateral MTG, ITG, and FG, left STG, IOG, LING, and cuneus, and right MOG, LN, and cerebellum. Although visual stimuli held an advantage, vocal emotions had their uniqueness, activating bilateral STG, TTG, PG, insula, and right MTG.

Prominent Dynamic Faces Compared to dynamic bodies, dynamic faces primarily activated right cerebellum, FG, PHG, and PG. Van de Riet et al. (2009) asked participants to categorize static facial and bodily expressions and found that faces activated cerebellum more than bodies; the current study extends this conclusion to dynamic faces and bodies. As is well known, the FG is a core region for face recognition, processing faces specifically. Atkinson et al. (2012) compared dynamic point-light displays of faces and bodies and found that dynamic faces significantly activated bilateral fusiform face areas compared to dynamic bodies. Another study (Kret et al., 2011) compared realistic threatening dynamic faces and bodies, showing that dynamic faces significantly activated the hippocampus compared to dynamic bodies. The PG is associated with action observation and imitation and is considered part of the human mirror neuron system (Iacoboni et al., 1999). This system provides an action recognition mechanism through imitation and learning, transforming sensory representations of actions into corresponding motor programs (Rizzolatti & Craighero, 2004). Numerous dynamic facial expression studies have found strong PG responses to dynamic faces, activated both in comparisons between dynamic and static faces (Foley et al., 2012; Sato et al.,

2012) and between dynamic emotional and neutral faces (Trautmann et al., 2009; Wicker et al., 2003). However, similar findings are rare in dynamic bodily expression studies.

Compared to vocal emotions, dynamic faces were more active in bilateral PHG, MTG, and cerebellum, left IFG and amygdala, and right STG, ITG, MOG, IOG, SFG, FG, PG, and IPL. Zhang Fan (2018) found that lower-level visual cortex and face-selective areas (FG and STG) showed greater activation for face pictures than vocal emotions. Other studies using dynamic stimuli found stronger activation for faces than voices in bilateral MTG and left ITG (Hasan et al., 2016). Aubé et al. (2015) added musical emotion to nonverbal vocalizations and found that dynamic faces activated occipital cortex, motor cortex, supplementary motor area (SMA), and dorsolateral prefrontal cortex (DLPFC). Meta-analyses have shown that dynamic faces activated bilateral PHG, amygdala, FG, visual association areas, left subcallosal gyrus, and right putamen more than vocal emotions (Dricu & Frühholz, 2016; Schirmer, 2018). A recent study (Lin et al., 2020) yielded richer results: dynamic faces showed greater activity than vocal emotions in bilateral MFG, left FG, IOG, cingulate gyrus, and caudate nucleus, and right ITG, postcentral gyrus, PG, supramarginal gyrus (SG), MOG, cuneus, and anterior cingulate gyrus. In summary, the current study validates previous findings and additionally discovers bilateral cerebellum and left IFG activation. The cerebellum primarily assists the brain in maintaining balance and coordination, playing an important role in sensory discrimination (Peng Danling, 2019). Therefore, it is unsurprising that faces, as part of the body, show greater cerebellar activity than vocal stimuli. Compared to voices, dynamic faces may generate more information requiring coordination, necessitating top-down cognitive control, thus activating IFG regions related to cognitive processes (Arsalidou et al., 2011; Christoff & Gabrieli, 2000).

Important Dynamic Bodies Contrast analysis results showed that dynamic bodies were more active than dynamic faces under dynamic, emotion, positive emotion, and negative emotion effects, involving left STG, IOG, LING, cuneus, and right MOG and cerebellum. Previous studies (Atkinson et al., 2012; Kret et al., 2011) have identified brain regions where dynamic bodies hold an advantage over dynamic faces, such as STG, cuneus, LING, and EBA, but did not find cerebellar activity. As early as 2001, Downing et al. discovered that EBA was sensitive to body stimuli but not to faces or other stimuli. The STG integrates complex human action information (de Gelder et al., 2014; Moro et al., 2008). The cuneus and LING primarily participate in emotion processing (Jastorff et al., 2015; Kret et al., 2011). The cerebellum is mainly responsible for visual perception and action understanding of body movements (Sokolov et al., 2010). The reason why cerebellum is activated in both dynamic face > dynamic body and dynamic body > dynamic face contrasts requires further investigation. Our preliminary speculation is that the cerebellum is connected to neural networks for both dynamic facial and bodily expressions, playing roles in both emotion and motion processing, but different cerebellar regions may be responsible for faces and bodies respectively. After all, the cerebellum is so large [a recent PNAS study found that the human cerebellum has nearly 80% of the neocortical surface area (Sereno et al., 2020)] and complex [divided into 13 lobules (Stoodley & Schmahmann, 2010) and numerous structures such as cerebellar peduncles and dentate nucleus (DN)], and no study has yet comprehensively parcellated the cerebellum anatomically and investigated specific functions of subregions.

Compared to vocal emotions, dynamic bodies significantly activated bilateral MTG, ITG, and FG, left STG, cuneus, LING, right MOG, and LN. This indicates that EBA not only distinguishes bodily expressions from other visual stimuli but also activates when discriminating vocal emotions. Additionally, the function of the fusiform body area (FBA) appears extended, not only processing body and emotion information (Goldberg, 2015; Hodzic et al., 2009) but also possibly participating in audiovisual information discrimination. Results show that cuneus and LING were only more strongly activated in dynamic bodies, not in dynamic faces or vocal emotions, suggesting that cuneus and LING are not only emotion-processing regions but may also be new specific brain regions for bodily expression processing. Although previous bodily expression studies did not report LN activation, they reported activation in its component, the putamen (de Gelder et al., 2004; van de Riet et al., 2009), which mainly processes motor and behavioral information. These findings provide a foundation for future research but require further empirical verification. To our knowledge, no empirical study has directly compared dynamic bodies and vocal emotions to investigate similarities and differences in their neural mechanisms.

Unique Vocal Emotions Compared to dynamic faces, vocal emotions activated bilateral STG and insula and left TTG more strongly. As previously mentioned, bilateral STG contains TVA, which specifically processes vocal emotions, consistent with other researchers' findings (Aubé et al., 2015; Schirmer, 2018). The insula primarily performs cognitive representation and emotional detection of vocal emotions, having been found activated in previous meta-analyses (Frühholz et al., 2016; Witteman et al., 2012) and in a recent empirical comparison between vocal emotions and dynamic faces (Lin et al., 2020). The TTG mainly processes identity information at the perceptual level and is an important component of the integrative model (Maguinness et al., 2018). Compared to dynamic bodies, vocal emotions activated bilateral STG, TTG, PG, left insula, and right MTG. These findings collectively demonstrate that TVA is an important specific brain structure for auditory information processing, effectively distinguishing vocal emotions from dynamic faces and bodies.

Additionally, beyond their inherent functions, TTG and insula may be further specific brain regions for vocal emotion processing. MTG also plays important roles in emotion processing and identity representation (Friederici & Alter, 2004; Roswandowitz et al., 2017). As part of frontal emotion-processing regions, PG plays an important role in emotion processing (Brattico et al., 2011; Fecteau et al., 2007; Numminen-Kontti, 2014).

This study has several limitations. First, the age range of included participants was not strictly restricted. Although previous ALE meta-analyses have included small numbers of children/adolescents or older adults (Arsalidou et al., 2020; Ran & Zhang, 2018), age is an important factor that may introduce inconsistent results. Future studies should restrict participant age ranges, conduct meta-analyses for specific age groups, and compare similarities and differences across age groups. Second, due to the limited number of relevant studies, insufficient for meeting basic ALE meta-analysis requirements for experimental contrasts (Eickhoff et al., 2017; Eickhoff et al., 2009), analysis of specific emotion types (e.g., anger, fear, disgust) was not conducted. Future meta-analyses should investigate neural mechanisms of each emotion type when sufficient data are available, and analyze similarities and differences in neural activity for the same emotion type across different carriers.

In conclusion, our findings validate and extend existing neural models of the three emotion carriers, clarifying similarities and differences in their neural mechanisms. Three separate meta-analyses revealed that dynamic facial expressions engage extensive frontal, occipital, temporal, and partial parietal cortices, as well as subcortical regions including hippocampus, cerebellum, thalamus, and amygdala; dynamic bodily activation concentrates in temporo-occipital regions, cerebellum, and hippocampus; and vocal emotions activate temporal and frontal lobes, amygdala, and insula. Conjunction analysis showed that the three emotion carriers cross-modally activated left MTG and right STG. Contrast analysis demonstrated that visual stimuli hold an advantage over auditory stimuli, particularly dynamic facial expressions, though dynamic bodily expressions also play an important role, while vocal emotions have their own uniqueness. Future research should further validate and extend these findings, investigate neural mechanisms of emotion processing across different age groups and their similarities and differences, examine brain mechanisms of specific emotion types and similarities/differences in neural activity for the same emotion across carriers, explore connectivity between different brain regions and different functions within the same region, and pay special attention to the neural basis of dynamic bodily expressions.

*Indicates literature included in the meta-analysis

Cao Linjing. (2018). *Research on cross-modal emotional representation mechanisms based on visual and auditory activation patterns* (Master' s thesis). Tianjin University.

Peng Danling. (Ed.). (2019). *General Psychology.* Beijing Normal University Press.

Zhang Fan. (2018). *Similarities and differences in neural mechanisms of emotional perception across visual and auditory modalities* (Master' s thesis). Southwest University, Chongqing.

Zhang Qi, Yin Tianzi, & Ran Guangming. (2015). Psychological mechanisms and neural basis of the dynamic facial expression advantage effect. *Advances in*

Psychological Science, 23(9), 1514–1525.

Zhou Linshu, Zhao Huaiyang, & Jiang Cunmei. (2017). Effects of music performance training on neuroplasticity: A meta-analytic study. *Advances in Psychological Science*, 25(11), 1869–1879.

Alba-Ferrara, L., Hausmann, M., Mitchell, R. L., & Weis, S. (2011). The neural correlates of emotional prosody comprehension: Disentangling simple from complex emotion. PLoS ONE*, 6(12), e28701. https://doi.org/10.1371/journal.pone.0028701

Arsalidou, M., Morris, D., & Taylor, M. J. (2011). Converging evidence for the advantage of dynamic facial expressions. Brain Topography*, 24(2), 149–163.

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: Brain responses to music, vocalizations and facial expressions. Social Cognitive and Affective Neuroscience*, 10(3), 399–407.

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., & Tervaniemi, M. (2011). A functional MRI study of happy and sad emotions in music with and without lyrics. Frontiers in Psychology*, 2, 308. https://doi.org/10.3389/fpsyg.2011.00308

Ceravolo, L. (2013). Neuroimaging and spatio-temporal dynamics of the localization and spatial attentional modulation by angry prosody* (Unpublished doctoral dissertation). Université de Genève.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. Neuroimage*, 36(2), 480–487.

Foley, E., Rippon, G., Thai, N. J., Longe, O., & Senior, C. (2012). Dynamic facial expressions evoke distinct activation in the face perception network: A connectivity analysis study. Journal of Cognitive Neuroscience*, 24(2), 507–520.

Grèzes, J., Pichon, S., & de Gelder, B. (2007). Perceiving fear in dynamic body expressions. NeuroImage*, 35(2), 959–966.

Jastorff, J., Huang, Y. A., Giese, M. A., & Vandenbulcke, M. (2015). Common neural correlates of emotion perception in humans. Human Brain Mapping*, 36(10), 4184–4201.

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. NeuroImage*, 18(1), 156–168.

Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., ···Gur, R. C. (2010). “It’ s not what you say, but how you say it” : A reciprocal temporo-frontal network for affective prosody. Frontiers in Human Neuroscience*, 4, 19. https://doi.org/10.3389/fnhum.2010.00019

Numminen-Kontti, T. (2014). Personality affects musical emotion processing: An fMRI-study* (Unpublished master’ s thesis). University of Helsinki.

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. Journal of Neuroscience*, 30(30), 10127-10134.

Poyo Solanas, M., Vaessen, M., & de Gelder, B. (2020). Computation-based feature representation of body expressions in the human brain. Cerebral Cortex*, 30(12), 6376-6390.

Regenbogen, C., Schneider, D. A., Gur, R. E., Schneider, F., Habel, U., & Kellermann, T. (2012). Multimodal human communication-targeting facial expressions, speech content and prosody. NeuroImage*, 60(4), 2346-2356.

Sato, W., Kochiyama, T., & Yoshikawa, S. (2010). Amygdala activity in response to forward versus backward dynamic facial expressions. Brain Research*, 1315, 92-99.

Sato, W., Toichi, M., Uono, S., & Kochiyama, T. (2012). Impaired social brain network for processing dynamic facial expressions in autism spectrum disorders. BMC Neuroscience*, 13(1), 1-17.

Trautmann, S. A., Fehr, T., & Herrmann, M. (2009). Emotions in motion: Dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. Brain Research*, 1284, 100-115.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. Neuron*, 40(3), 655-664.

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

Source: ChinaXiv —Machine translation. Verify with original.