NeuroImage 174 (2018) 1–10

Contents lists available at ScienceDirect

NeuroImage

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

Decoding the neural signatures of emotions expressed through sound

Matthew E. Sachs [*], Assal Habibi, Antonio Damasio, Jonas T. Kaplan

Brain and Creativity Institute, University of Southern California, 3620A McClintock Avenue, Los Angeles, CA, 90089-2921, United States

ARTICLE INFO

Keywords:
fMRI
Music
Voice
Emotions
Multivoxel pattern analysis

ABSTRACT

Effective social functioning relies in part on the ability to identify emotions from auditory stimuli and respond appropriately. Previous studies have uncovered brain regions engaged by the affective information conveyed by sound. But some of the acoustical properties of sounds that express certain emotions vary remarkably with the instrument used to produce them, for example the human voice or a violin. Do these brain regions respond in the same way to different emotions regardless of the sound source? To address this question, we had participants ($N = 38$, 20 females) listen to brief audio excerpts produced by the violin, clarinet, and human voice, each conveying one of three target emotions—happiness, sadness, and fear—while brain activity was measured with fMRI. We used multivoxel pattern analysis to test whether emotion-specific neural responses to the voice could predict emotion-specific neural responses to musical instruments and vice-versa. A whole-brain searchlight analysis revealed that patterns of activity within the primary and secondary auditory cortex, posterior insula, and parietal operculum were predictive of the affective content of sound both within and across instruments. Furthermore, classification accuracy within the anterior insula was correlated with behavioral measures of empathy. The findings suggest that these brain regions carry emotion-specific patterns that generalize across sounds with different acoustical properties. Also, individuals with greater empathic ability have more distinct neural patterns related to perceiving emotions. These results extend previous knowledge regarding how the human brain extracts emotional meaning from auditory stimuli and enables us to understand and connect with others effectively.

Introduction

The capacity to both convey and perceive emotions through sounds is crucial for successful social interaction. For example, recognizing that a person is distressed based on vocal expressions alone can confer certain advantages when it comes to communicating and connecting with others. Intriguingly, emotions can be recognized in non-vocal sounds as well. Music can convey emotions even when not mimicking the human voice, despite the fact that an ability to express emotions through music does not serve as clear an evolutionary function as vocal expressions of emotions (Frühholz et al., 2014). And yet, the capability to consistently and reliably discern musical emotions appears to be universal, even in individuals with no musical training (Fritz et al., 2009). Studying the neural overlap of expressions of emotions in both vocal and musical stimuli therefore furthers our understanding of how auditory information becomes emotionally relevant in the human brain.

Previous univariate neuroimaging studies that have examined this neural overlap have reported activity in the superior temporal gyrus (Escoffier et al., 2013), amygdala and hippocampus (Frühholz et al., 2014) during both musical and non-musical, vocal expressions of emotions. While these results support the notion that musical and vocal patterns recruit similar brain regions when conveying emotions, they do not clarify whether these regions are responsive to a specific emotional category or are involved in emotion processing more generally. Neither study addressed the neural activity patterns that are specific to a particular emotion, but conserved across the two different domains of music and vocals. One particular univariate study did attempt to answer this question, but only with the emotion of fear: the researchers found that the amygdala and posterior insula were commonly activated in response to fear expressed through non-linguistic vocalizations and musical excerpts, as well as through facial expressions, (Aubé et al., 2013).

In general, however, univariate methods are not well suited for evaluating commonalities in the processing of emotions across the senses because, due to spatial smoothing and statistical limitations, they cannot assess information that may be located in fine-grained patterns of activity

---

[*] Corresponding author.
E-mail address: msachs@usc.edu (M.E. Sachs).

https://doi.org/10.1016/j.neuroimage.2018.02.058
Received 11 October 2017; Received in revised form 23 February 2018; Accepted 27 February 2018
Available online 1 March 2018

dispersed throughout the brain (Kaplan et al., 2015). Multivoxel pattern analysis (MVPA), which entails classifying mental states using the spatially-distributed pattern of activity in multiple voxels at once, can provide a more sensitive measure of the brain regions that are responsible for distinguishing amongst different emotions (Norman et al., 2006). In combination with a searchlight analysis, in which classification is performed on local activity patterns within a sphere that traverses the entire brain volume, MVPA can reveal areas of the brain that contain information regarding emotional categories (Kriegeskorte et al., 2006; Peelen et al., 2010). This multivariate approach has been used in various capacities to predict emotional states from brain data (Saarimaki et al., 2015). Spatial patterns within the auditory cortex, for example, were used to classify emotions conveyed through both verbal (Ethofer et al., 2009) and nonverbal (Kotz et al., 2013) speech. However, it remains unclear whether the neural activity in these regions correspond to a particular category of emotion or are instead only sensitive to the lower-level acoustical features of sounds.

Multivariate cross-classification, in which a classifier is trained on brain data corresponding to an emotion presented in one domain and tested on separate brain data corresponding to an emotion presented in another, is a useful approach to uncovering representations that are modality independent (see Kaplan et al., 2015 for review). Previously, this approach has been used to demonstrate that emotions induced by films, music, imagery, facial expressions, and bodily actions can be successfully classified across different sensory domains (Peelen et al., 2010; Skerry and Saxe, 2014; Kragel and LaBar, 2015; Saarimaki et al., 2015; Kim et al., 2017). Cross-modal searchlight analyses revealed that successful classification of emotions across the senses and across sources could be achieved based on signal recorded from the cortex lying within the superior temporal sulcus (STS), the posterior insula, the medial prefrontal cortex (MPFC), the precuneus, and the posterior cingulate cortex (Kim et al., 2010; Peelen et al., 2010; Saarimaki et al., 2015). While informative for uncovering regions of the brain responsible for representing emotions across the senses, these studies did not address how the brain represents emotions within a single sensory domain when expressed in different ways. To our knowledge, there has been no existing research on the affect-related neural patterns that are conserved across vocal and musical instruments, two types of auditory stimuli with differing acoustical properties.

Additionally, the degree to which emotion-specific predictive information in the brain might be modulated by individual differences remains unexplored. Empathy, for example, which entails understanding and experiencing the emotional states of others, is believed to rely on the ability to internally simulate perceived emotions (Lamm et al., 2007). Activation of the anterior insula appears to be related to linking observed expressions of emotions with internal empathic responses (Carr et al., 2003) and the degree of activation during emotion processing tasks is shown to be positively correlated with measures of empathy (Singer et al., 2004; Silani et al., 2008). Emotion-distinguishing activity patterns in the insula may therefore relate to individual differences in the tendency to share in the affective states of others.

Here, we used MVPA and cross-classification on two validated datasets of affective auditory stimuli, one of non-verbal vocalizations (Belin et al., 2008) and one of musical instruments (Paquette et al., 2013), to determine if patterns of brain activity can distinguish discrete emotions when expressed through different sounds. Participants were scanned while listening to brief (0–4s) audio excerpts produced by the violin, clarinet, and human voice and designed to convey one of three target emotions—happiness, sadness, and fear. The authors who published the original dataset chose the violin and clarinet because both musical instruments can readily imitate the sounds of the human voice, but are from two different classes (strings and woodwinds respectively; Paquette et al., 2013). These three target emotions were used because (1) they constitute what are known as “basic” emotions, which are believed to be universal and utilitarian (Ekman, 1992), (2) they can be reliably produced and conveyed on the violin and clarinet (Hailstone et al., 2009) and (3) they are also present in both the vocal and musical datasets.

After scanning, a classifier was trained to differentiate the spatial patterns of neural activity corresponding to each emotion both within and across instruments. To understand the contribution of certain acoustic features to our classification results, we compared cross-instrument classification accuracy with fMRI data to cross-instrument classification accuracy using acoustic features of the sounds alone. Then, a searchlight analysis was used to uncover brain areas that represent the affective content that is shared across the two modalities, i.e. music and the human voice. Finally, classification accuracies within a priori-defined regions of interest in the auditory cortex, including the superior temporal gyrus and sulcus, as well as the insula were correlated with behavioral measures of empathy. These regions were selected for further investigation because of their well-validated roles in the processing of emotions from sounds (Bamiou et al., 2003; Sander and Scheich, 2005) as well as across sensory modalities (Peelen et al., 2010; Saarimaki et al., 2015). Based on previous results, we predict that BOLD signal in the auditory and insular cortices will yield successful classification of emotions across all three instruments. Moreover, given the known role of the insula in internal representations of observed emotional states (Carr et al., 2003), we hypothesize that classification accuracies within the insula will be positively correlated with empathy.

Materials and methods

Participants

Thirty-eight healthy adult participants (20 females, mean age = 20.63, SD = 2.26, range = 18–31) were recruited from the University of Southern California and surrounding Los Angeles community. All participants were right-handed, had normal hearing and normal or corrected-to-normal vision, and had no history of neurological or psychiatric disorders. All experimental procedures were approved by the USC Institutional Review Board. All participants gave informed consent and were monetarily compensated for participating in the study.

Survey

The Goldsmith Musical Sophistication Index (Gold-MSI; Mullensiefen, et al., 2014) was used to evaluate past musical experience and degree of music training. The Gold-MSI contains 39 items broken up into five subscales, each related to a separate component of musical expertise: *active engagement, perceptual abilities, musical training, singing abilities,* and *emotions.* The scale also contains a *general musical sophistication* score, which is the sum of responses to all items. Each item is scored on a 7-point Likert scale from 1 = *completely disagree* to 7 = *completely agree.*

Both cognitive and affective components of empathy were measured using the Interpersonal Reactivity Index (Davis, 1983), which includes 28 items and four subscales: fantasy and perspective taking (cognitive empathy) and empathic concern and personal distress (affective empathy). Supplementary Table 1 summarizes the results obtained from the surveys.

Stimuli

Two validated, publically-available datasets of short, affective auditory stimuli were used: the Music Emotional Bursts (MEB; Paquette et al., 2013) and the Montreal Affective Voices (MAV; Belin et al., 2008). Studying neural responses to relatively short stimuli provided two main advantages: (1) As suggested Paquette et al. (2013), these brief bursts of auditory emotions may mimic more primitive, and therefore more biologically relevant, expressions of affect and (2) it allows us to maximize the number of trials that can be presented to participants in the scanner, theoretically improving the training of the classifier. The MEB contains 60 brief (1.64s on average) auditory clips designed to express 3 basic emotions (happiness, sadness, and fear) played on either the violin or