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

clarinet. The dataset contains 10 unique exemplars of each emotion on each instrument. The MAV is a set of brief (1.35s on average) non-verbal vocalizations that reliably convey the same 3 emotions (happiness, sadness, and fear) as well as several others (disgust, pain, surprise, and pleasure; these emotional clips were not included in this study because they were not included in the MEB dataset). The MAV dataset contains 10 unique exemplars of each emotion as well and includes both female and male voices. Both the MEB and MAV also include a neutral condition, which were not included in this study either. All clips from both datasets had been normalized so that the peak signal value corresponded to 90% of the maximum amplitude (Belin et al., 2008; Paquette et al., 2013). Combining these two stimulus datasets resulted in 90 unique stimuli: 30 featuring the violin, 30 featuring the clarinet, and 30 featuring the human voice, with 30 clips for each of the three emotions (happiness, sadness, and fear).

Design and procedure

Stimuli were presented in an event-related design using MATLAB's PsychToolbox (Kleiner et al., 2007) in 6 functional runs. During scanning, participants were instructed to be still, with their eyes open and focused on a fixation point continually presented on a screen, and attend to the audio clips when they heard them. Auditory stimuli were presented through MR-compatible OPTOACTIVE headphones with noise-cancellation (Optoacoustics). An eye-tracking camera was monitored to ensure that the participants were awake and alert during scanning.

During each functional run, participants listened to 45 audio clips, 5 clips for each trial type (emotion x instrument). Each clip was followed by a rest-period that varied in length and resulted in a total event length time (clip + rest) of 5s, regardless of the length of the clip. Five, 5-s rest events, in which no sound played, were also added as an additional condition, resulting in a total functional run time of 250s (125 TRs, see Fig. 1). Two unique orders of stimuli presentation were created using a genetic algorithm (Kao et al., 2009), which takes into account designed detection power and counterbalancing to generate an optimal design that is pseudorandomized. One optimized order of stimuli presentation was used on odd-numbered runs (1, 3 and 5) and the other order was used on even-numbered runs (2, 4, and 6). Over the course of the 6 functional runs, each of the 90 audio stimuli were presented exactly 3 times.

To validate the accuracy of the clips in terms of their ability to convey the intended emotion, after scanning, participants listened to all 90 clips again in random order and selected the single emotion, from the list of three, that they believed was being expressed in the clip. To further describe their perceptions of each clip, participants also rated each clip for how intensely it expressed each of the three target emotions using a scale ranging from 1 (not at all) to 5 (very much).

Data acquisition

Images were acquired with a 3-T Siemens MAGNETON Prisma System and using a 32-channel head coil. Echo-planar volumes were acquired continuously with the following parameters: repetition time (TR) = 2000 ms, echo time (TE) = 25 ms, flip angle = 90°, 64 × 64 matrix, in-plane resolution 3.0 × 3.0 mm, 41 transverse slices, each 3.0 mm thick, covering the whole brain. Structural T1-weighted magnetization-prepared rapid gradient echo (MPRAGE) images were acquired with the following parameters: TR = 2530 ms, TE = 3.09 ms, flip angle = 10°, 256 × 256 matrix, 208 coronal slices, 1 mm isotropic resolution.

Data processing

Data preprocessing and univariate analysis was done in FSL (FMRIB Software Library, Smith et al., 2004). Data were first pre-processed using brain extraction, slice-time correction, motion correction, spatial smoothing with 5 mm FWHM Gaussian kernel, and high-pass temporal filtering. Each of the 9 trial types (emotion*instrument) was modeled with a separate regressor derived from a convolution of the task design and a double gamma hemodynamic response function. Six motion correction parameters were included in the design as nuissance regressors. The functional data were registered to the high-resolution anatomical image of each subject and to the standard Montreal Neurological Institute (MNI) brain using the FSL FLIRT tool (Jenkinson and Smith, 2001). Functional images were aligned to the high-resolution anatomical image using a 7 degree-of-freedom linear transformation. Anatomical images were registered to the MNI-152 brain using a 12 degree-of-freedom affine transformation. This entire procedure resulted in one statistical image for each of the 9 trial types (3 emotions by 3 instrument) in each run. Z-stat images were then aligned to the first functional run of that participant for within-subject analysis.

Multivoxel pattern analysis

Multivoxel pattern analysis (MVPA) was conducted using the PyMVPA toolbox (http://www.pymvpa.org/) in Python. A linear support vector machine (SVM) classifier was trained to classify the emotion of each trial type. Leave-one run out cross-validation was used to evaluate classification performance (i.e. 6-fold cross-validation with 45 data points in the training dataset and 9 data points in the testing dataset for each fold). Classification was conducted both within each instrument as well as across instruments (training the classifier on a subset of data from two of the instruments and testing on a left-out subset from another instrument) using a mask of the participant's entire brain. In addition to training on two instruments and testing on the third, we also ran cross-instrument classification for every pairwise combination of training on one instrument and testing on another (6 combinations in total). Feature selection on the whole brain mask was employed on the training data alone using a one-way ANOVA and keeping the top 5% most informative voxels (mean 3320 voxels after feature selection, SD = 251). Within participant classification accuracy was computed by averaging the accuracy of predicting the emotion across each of the 6 folds. One-sample t-tests on the population of participant accuracies were performed to determine if the achieved accuracies were significantly above theoretical chance (33%).

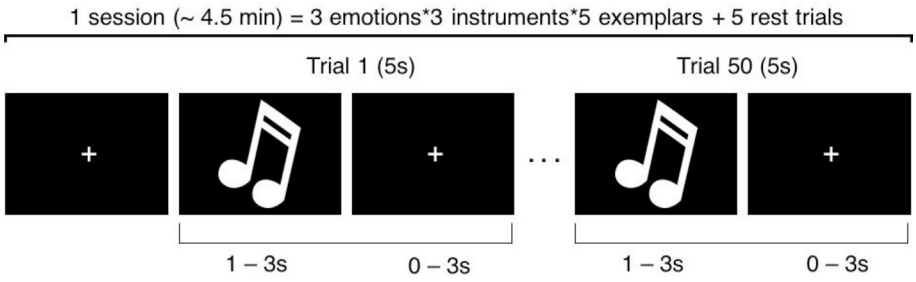

Fig. 1. Example of one functional session. In each session, participants listened to 45 clips and 5 rest trials, each of which lasted for a total of 5s. Each functional session lasted around 4.5min and there were six functional scans in total.

Region of interest classification

In addition to whole brain analysis, we performed a region of interest (ROI) analysis focusing on a-priori ROIs in the auditory cortex and insular cortex. These two ROIs were chosen because of their well-known roles in the processing of emotions from sounds (Bamiou et al., 2003; Sander and Scheich, 2005). For the auditory cortex, we used the Harvard-Oxford Atlas planum temporale mask, which is defined as the superior surface of the superior temporal gyrus, as well as the Heschl's gyrus mask, merged and thresholded at 25 (Fig. 2A). For the insula, we used masks of the dorsal anterior, ventral anterior, and posterior insula described in Deen et al. (2011) that were defined by the results of a cluster analysis of functional connectivity patterns (Fig. 2B). Within and across instrument classification was conducted in exactly the same way as described above. For the region of interest analysis, feature selection was not used, that is, all voxels within the specified anatomical region were used.

Whole-brain searchlight analysis

A searchlight analysis for classifying emotions was conducted both within and across modalities (Kriegeskorte et al., 2006). For each subject, the classification accuracy was determined for spheres with radius 3 voxels throughout the entire brain. A sphere of that size was chosen to roughly match the size of the anatomical regions of interest, large enough to not be biased by individual variation in any one voxel and yet small enough to adhere to known anatomical boundaries. These accuracies were then mapped to the center voxel of the sphere and warped to standard space. The searchlight analysis was conducted both within instruments and across instruments. For the within instrument searchlights, the SVM classifier was trained on data from all but one of the six runs and tested on the left-out run (leave-one-run-out cross validation). To evaluate the significance of clusters in the overlapped searchlight accuracy maps, nonparametric permutation testing was performed using FSL's Randomise tool (Winkler et al., 2014), which models a null distribution of expected accuracies at chance. The searchlight accuracy maps were thresholded using threshold-free cluster enhancement (TFCE; Smith and Nichols, 2009).

For the cross-instrument searchlight analysis, the classifier was trained on data from every combination of two instruments and tested on data from the left-out instrument, resulting in a total of three cross-instrument accuracy maps. The three cross-instrument searchlights were also overlayed to determine the regions of overlap. To determine the significance of cross-classification searchlight, we used a more complex nonparametric method than what was used to determine significance of the within-instrument searchlight maps. As described in Stelzer et al. (2013), this method involves random permutation tests on the subject level combined with bootstrapping at the group level. While Randomise with TFCE, which was used to determine significance of the within-instrument searchlight maps, does provide excellent control of type 1 errors, the Stelzer et al. (2013) can provide a more accurate estimation of the group level statistics because it models the null distribution of searchlight maps on both the individual subject level and group level. However, because within-subject permutation testing and across-subject bootstrapping is computationally intensive, we only used this method for determining significance thresholds for the cross-modality searchlight maps, not for the within-instrument searchlights maps. We believe this decision is justified because a) within modality classification in auditory cortex is already well known and does not require a higher standard of proof, b) successful cross-modal classification implies successful within modality classification, and c) the cross-modal searchlights constitute the most direct test of our hypotheses.

To achieve this, we randomly permuted the class labels 50 times and performed whole-brain cross searchlight analyses to create 50 single subject chance accuracy maps. One permuted accuracy map per subject was selected at random (with replacement) to create a pooled group accuracy map. This procedure was repeated 10,000 times to create a distribution of pooled group accuracy maps. Next, a threshold accuracy was found for each voxel by determining the accuracy that corresponded to a p-value of 0.001 in the voxel-wise pooled group accuracy map. Clusters were then defined as a group of contiguous voxels that survived these voxel-wise accuracy thresholds and cluster sizes were recorded for each of the 10,000 permuted group accuracy maps to create a histogram of cluster sizes at chance. Finally, cluster-sizes from the chance

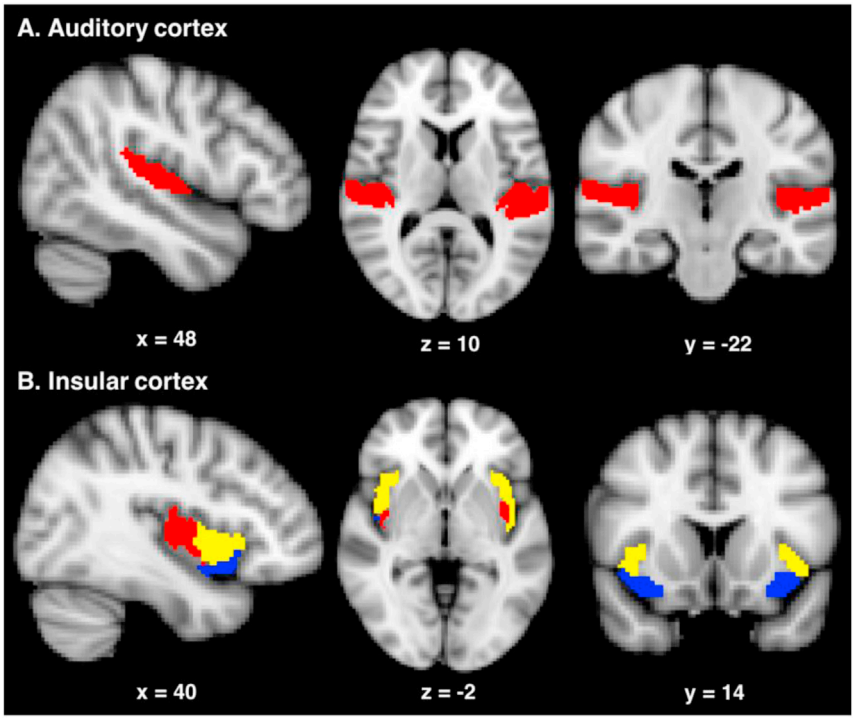

Fig. 2. Regions of interest. ***A,*** The auditory cortex was defined using the Harvard-Oxford Atlas by merging the planum temporale mask with Heschl's gyrus mask, both thresholded at 25. ***B,*** Three major subdivisions, the dorsal anterior, ventral anterior, and posterior, were identified based on the results from a previous study using cluster analysis of functional connectivity patterns.

distribution were compared to cluster-sizes from the original, group accuracy maps to determine significance. An FDR-method using Benjamini-Hochberg procedure was used to correct for multiple comparisons at the cluster level (Heller et al., 2006).

Multiple regression with personality measures

Individual scores on the empathy subscales of the IRI were correlated with classification accuracy within the four ROIs for both within and across classification to determine if the degree of emotion-specific predictive information within these regions is associated with greater emotional empathy. Age, gender, and music sophistication, as measured by the Gold-MSI, were included in the model as regressors of no interest. Additionally, behavioral accuracy of correctly identifying the intended emotions of the sound clips collected outside of the scanner were correlated with performance of the classifier.

Acoustic features of sound clips

For extracting acoustic features from the sound clips believed to be relevant to emotional expression, we used MIRToolbox, a publically available MATLAB toolbox primarily used for music information retrieval (Lartillot and Toiviainen, 2007), but well suited for extracting relevant acoustical information from non-musical and vocal stimuli as well (Linke and Cusack, 2015; Rigoulot et al., 2015). These included: spectral centroid, spectral brightness, spectral flux, spectral rolloff, spectral entropy, spectral spread, and spectral flatness for evaluating timbral characteristics, RMS energy for evaluating dynamics, mode and key clarity for evaluating tonal characteristics, and fluctuation entropy and fluctuation centroid for evaluating rhythmic characteristics of the clips (Alluri et al., 2012). We additionally added the acoustic features published in Paquette et al. (2013), which included duration, mean fundamental frequency, max fundamental frequency, and min fundamental frequency. We then evaluated how these features varied by instrument and by emotion and conducted a classification analysis based on acoustic features alone to predict the intended emotion of the sound clip.

Because we found a main effect of emotion label on duration of the clips (i.e. fear clips were significantly shorter than sad clips) and we do not believe that this different reflects a meaningful difference amongst emotions, we added an additional regressor of no interest where the height of the regressor reflected the duration of each clip in a separate GLM analysis. MVPA and searchlight analysis were then repeated with this model for comparison.

Results

Behavioral results

Behavioral ratings of the sound clips outside of the scanner were collected for 37 out of the 38 participants. Overall, participants correctly labeled 85% of the clips (SD = 17%). Averaged correct responses for each emotion and instrument are presented in Table 1. The between-within ANOVA on accuracy scores for each of the clips showed a significant interaction between emotion and instrument ($F(4,144) = 64.30$, $p < 0.0001$). Post-hoc follow-up t-tests showed that the fear condition on the clarinet was more consistently labeled incorrectly (mean accuracy = 55%, SD = 16%) than the happy condition on the clarinet (mean accuracy = 97%, SD = 5%; $t(36) = 15.99$, $p < 0.0001$, paired *t*-test) as well as the fear condition in the violin (mean accuracy = 90%, SD = 13%; $t(36) = 13.83$, $p < 0.0001$, paired *t*-test) or voice (mean accuracy = 94%, SD = 7%; $t(36) = 13.27$, $p < 0.0001$, paired *t*-test).

For intensity ratings of the clips, we calculated the average intensity of each emotion for each participant. Again, an interaction between emotion and instrument was found for the intensity ratings ($F(4,144) = 38.73$, $p < 0.0001$, ANOVA). Fear clips on the clarinet were rated as significantly less intense than fear clips on the violin ($t(36) = 14.49$, $p < 0.0001$, paired *t*-test) and voice ($t(36) = 13.16$, $p < 0.0001$, paired *t*-test), whereas sad clips on the voice were rated as significantly more intense than sad clips on the violin ($t(36) = 9.84$, $p < 0.0001$, paired *t*-test) or clarinet ($t(36) = 9.90$, $p < 0.0001$, paired *t*-test; see Table 1 for average ratings of intensity and average accuracy). Overall, the intensity ratings provide further information for the accuracy scores: fear on the clarinet was the most difficult to identify and was rated as significantly less intense. Participant intensity ratings were not found to be related to the performance of the brain-based classifier.

Multivariate results

MVPA applied to the whole brain to predict the emotion of each clip showed above chance (0.33) accuracies using data from all instruments ($M = 0.43$, $SD = 0.08$, $t(37) = 7.58$, $p < 0.0001$). Above chance accuracy was also obtained using data collected from each instrument individually (*clarinet:* $M = 0.39$, $SD = 0.12$, $t(37) = 3.06$, $p = 0.004$; *violin*: $M = 0.37$, $SD = 0.11$, $t(37) = 2.18$, $p = 0.04$; *voice*: $M = 0.43$, $SD = 0.15$, $t(37) = 4.14$, $p = 0.0002$). Within instrument classification accuracy was also significantly above chance in both the auditory cortex ($M = 0.49$, $SD = 0.06$, $t(37) = 15.06$, $p < 0.0001$) and all three regions of the insula (*dorsal anterior:* $M = 0.38$, $SD = 0.07$, $t(37) = 4.06$, $p = 0.0002$; *ventral anterior:* $M = 0.38$, $SD = 0.07$, $t(37) = 4.41$, $p < 0.0001$; *posterior:* $M = 0.38$, $SD = 0.08$, $t(37) = 3.68$, $p = 0.0007$, see Fig. 3). Confusion matrices for within instrument classification in the whole brain and auditory cortex are provided in the supplementary materials (Supplementary Fig. 1) as well as additional measures of classification performance, including sensitivity, specificity, positive predictive value, and negative predictive value (Supplementary Table 2).

Cross-classification accuracies, in which the classifier was trained on data from two instruments and tested on data from the left-out third instrument, also showed successful classification for each combination of training and testing (3 in total). Classification accuracy averaged across the 3 combinations of training and testing was significantly greater than chance in the whole brain ($M = 0.38$, $SD = 0.09$, $t(37) = 3.56$, $p = 0.001$) as well as the region of interest in the auditory cortex ($M = 0.44$, $SD = 0.08$, $t(37) = 8.83$, $p < 0.0001$), but not in the three insula ROIs. Graphs of the accuracies for each combination of training and testing in both the whole brain analysis and ROI analysis are presented in Fig. 3. Confusion matrices for cross instrument classification in the whole brain and auditory cortex are provided in Supplementary Fig. 2.

We additionally conducted cross-classification for all 6, pairwise combinations of training on one instrument and testing on one other instrument. Overall, the pairwise cross-instrument classification

Table 1
Behavioral ratings of intensity and emotion label for stimuli by instrument and emotion. Accuracy is calculated as the number of clips correctly identified with the intended emotion label.

	Happy		Sad		Fear		Total	
	Acc	Intensity	Acc	Intensity	Acc	Intensity	Acc	Intensity
All	0.92	3.86 (0.58)	0.85	3.81 (0.68)	0.79	3.52 (0.87)	0.85	3.86 (0.58)
Voice	0.82	4.17 (0.45)	0.81	4.35 (0.48)	0.90	4.04 (0.56)	0.84	4.19 (0.51)
Violin	0.96	3.54 (0.61)	0.89	3.54 (0.56)	0.94	3.87 (0.69)	0.93	3.64 (0.64)
Clarinet	0.97	3.85 (0.50)	0.85	3.55 (0.64)	0.55	2.67 (0.61)	0.79	3.36 (0.77)

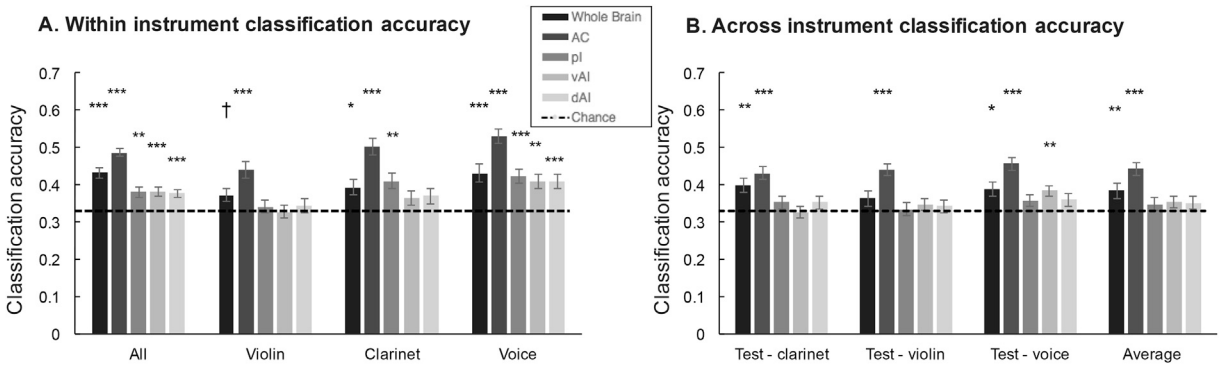

Fig. 3. Classification accuracies for MVPA decoding of emotions in auditory stimuli using fMRI data from the whole brain and four region of interest. ***A***, Classification accuracies in the whole brain, auditory cortex (AC), posterior insula (pI), dorsal anterior insula (dAI), and ventral anterior insula (vAI) with all three instruments (violin, clarinet, voice) as well as within each instrument individually. ***B***, Cross-instrument classification accuracies in the whole brain, auditory cortex (AC), posterior insula (pI), dorsal anterior insula (dAI), and ventral anterior insula (vAI), leaving out data from one instrument and training on the other two. Error bars represent indicate error. p values are calculated based on a one-sample *t*-test comparing classification with chance (0.33, dotted line). †$p < 0.05$, uncorrected; *$p < 0.05$; **$p < 0.01$,***$p < 0.001$, corrected for multiple comparisons across the four ROIs.

accuracies were significantly above chance in the auditory cortex. The results are presented in Supplementary Fig. 3.

Searchlight results

The whole-brain, within instrument searchlight analysis revealed that successful classification of the emotions of the musical clips could be found bilaterally in the primary and secondary auditory cortices, including the cortices lying within the superior temporal gyrus and sulcus, as well as the bilateral posterior insular cortices, parietal operculum, precentral gyrus, inferior frontal gyrus, right middle temporal gyrus, the right medial prefrontal cortex, right superior frontal gyrus, right precuneus, and right supramarginal gyrus (Fig. 4). Center coordinates and accuracies for significant regions in the within instrument searchlight analysis are presented in Supplementary Table 2.

Three whole-brain across instrument searchlight analyses were conducted where the classifier was trained on data from two of the instruments and tested on the held-out third instrument. All three searchlights showed significant classification bilaterally in primary auditory cortex, including Heschl's gyrus, and the superior temporal gyrus and sulcus, as well as the posterior insula and parietal operculum (Fig. 5). Several other brain regions showed significant classification in one or more of the searchlight analyses, but not all three. These included the right middle and inferior frontal gyri and precentral gyrus (leaving out the violin and the voice) and the MPFC (leaving out clarinet). Center coordinates and accuracies for each significant region of the cross-instrument searchlight analysis are presented in Supplementary Table 3.

Multiple regression results

Measures of the four subscales of the IRI were modeled in a multiple regression to predict the classification accuracies in each of the four regions of interest (auditory cortex and three subcomponents of the insula) with age and gender added as covariates of no interest. In this model, empathic concern was positively correlated with both the within and cross classification accuracies in the dorsal anterior insula (*Within:* $\beta = 0.08$, $p = 0.0101$, *Cross:* $\beta = 0.08$, $p = 0.0158$). The significance of the regression coefficient between empathic concern and within-instrument accuracy in the dorsal anterior insula survived correction for multiple comparisons, though the regression with cross-instrument accuracies did not (Bonferroni correction with four regions of interest, alpha = 0.0125). No other predictors were significantly correlated with accuracy in the regions of interest.

Scores corresponding to the five subscales of the MSI were additionally modelled in a separate multiple regression as covariates of interest to predict classification accuracies in the four ROIs. No significant correlations were found between musical experience and classification accuracy in either the auditory cortex or insula. Additionally, no significant correlations were found between behavioral accuracies of correctly identifying the intended emotion of the clip (collected outside of the scanner) and classification accuracy.

Acoustic features classification and duration

Duration was significantly different for the three emotions according to a one-way ANOVA ($F(2,87) = 110.30$, $p < 0.0001$). Sad clips ($M = 2.39$s, $SD = 0.54$) were significantly longer than the happy ($M = 1.48$s, $SD = 0.39$; $t(58) = 7.45$, $p_{adjust} = 1.2 \times 10^{-12}$) and fear clips ($M = 0.81$s, $SD = 0.25$; $t(58) = 14.38$, $p_{adjust} < 0.0001$). Because duration is not an acoustic feature directly related to the expression of an emotion, and because it differed significantly by emotion, we wanted to ensure that the classifier was not only classifying based on stimuli length rather than its emotional content. We therefore added the length of each clip as an additional parametric regressor in the lower-level GLM models and redid both within and cross instrument classification with the z-stat images obtained from this analysis. The average within instrument accuracy using duration as a regressor was 44% (SD = 7%) within the whole brain and the average across instrument accuracy using duration as a regressor

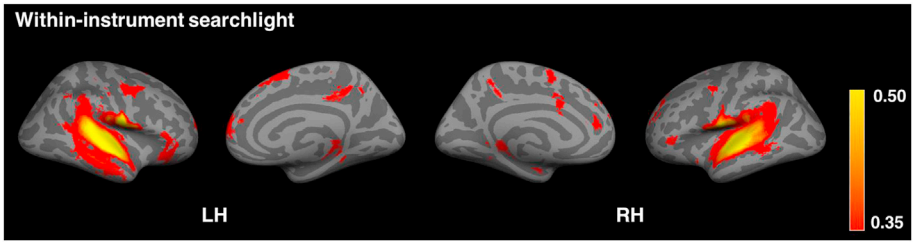

Fig. 4. Within instrument whole-brain searchlight results using data from all instruments and leave-one-run out cross validation. Red-yellow colors represent classification accuracy. Significant clusters determined by permutation testing. All images are thresholded to show clusters that reached a FDR-corrected significance level at alpha = 0.05.

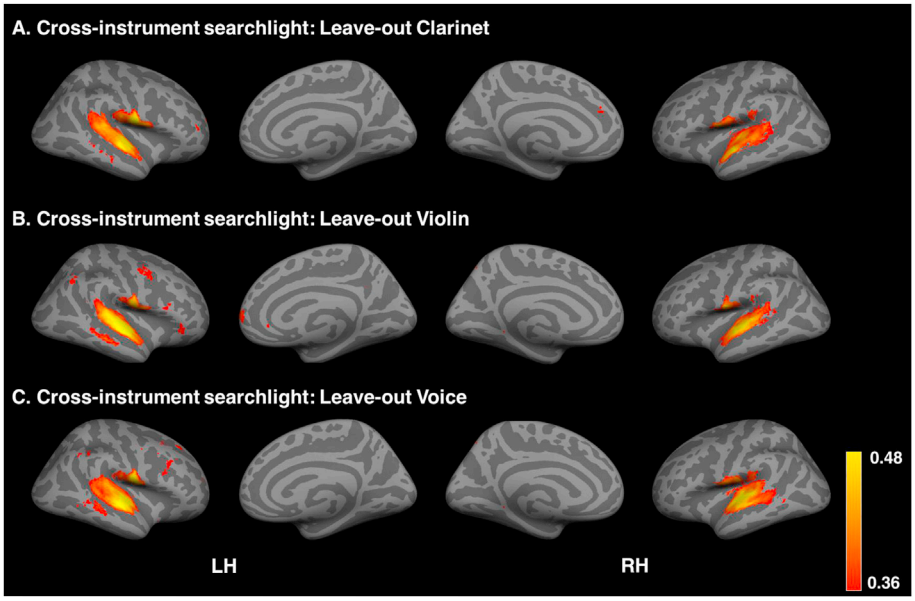

Fig. 5. *A,* Cross-instrument whole-brain searchlight results training on data collected during violin and voice clips, testing on clarinet clips. *B,* Cross-instrument whole-brain searchlight results training on data collected during clarinet and voice clips, testing on violin clips. *C,* Cross-instrument whole-brain searchlight results training on data collected during violin and clarinet clips, testing on voice clips. Red-yellow colors represent classification accuracy. Significant clusters were determined by permutation testing. All images are thresholded to show clusters that reached a FDR-corrected significance level at alpha = 0.05.

was 39% (SD = 9%), which were both statistically significant according to a one-way *t*-test against theoretical chance (*within:* $t(37) = 9.12$, $p = 5.33 \times 10^{-11}$; *across:* $t(37) = 4.22$, $p = 0.0002$). No significant differences were found between the classification accuracies when duration was added as a regressor (*within:* $t(37) = 0.61$, $p = 0.55$, *across:* $t(37) = 0.53$, $p = 0.60$, paired *t*-test). Because of these, we did not recompute the searchlight analysis using the results from the GLM analysis with duration modelled.

We also conducted a classification analysis using the acoustic features of the sound clips only. These included 12 features related to timbre, rhythm, and tonality as described in (Alluri et al., 2012) as well as fundamental frequency and duration (Paquette et al., 2013). The linear SVM classifier could successfully classify the emotion label of the sound clip 82% of the time using all data and 60% on average when training and testing on data from separate instruments (cross classification). The duration of the sound clips was determined to be the most important feature used by the SVM. When duration was removed, classification accuracy was 72% when using data from all instruments and 57% on average when training and testing across all three instruments (see Table 2). After removing duration, the most important features for classification were fluctuation centroid and spectral flux. Fluctuation centroid is a measure of rhythmic changes in sounds and is calculated by taking the mean (center of gravity) of the fluctuation spectrum, which conveys the periodicities contained in a sound wave's envelope (Alluri et al., 2012). A one-way ANOVA revealed a significant main effect of emotion $F(2,87) = 29.93$, $p < 0.0001$) on fluctuation centroid. Fear clips ($M = 2977$; $SD = 1220$) were significantly higher than both sad ($M = 1325$; $SD = 554$; $t(58) = 6.76$, $p_{adjust} < 0.0001$) and happy clips ($M = 2432$; $SD = 632.21$; $t(58) =$, $p_{adjust} < 0.0001$). Spectral flux is a measure of how the variance in the audio spectrum changes over time and therefore conveys both spatial and temporal components of sound (Alluri et al., 2012). It is highly correlated with fluctuation centroid with our sound clips. A one-way ANOVA revealed that fearful clips ($M = 175.03$, $SD = 86.05$) had a significantly higher spectral flux than both sad ($M = 59.70$, $SD = 33.32$; $t(58) = 6.85$, $p < 0.001$) and happy clips ($M = 93.20$, $SD = 28.25$; $t(58) = 4.95$, $p < 0.001$).

The results from the acoustic classification provide information regarding how the fMRI-based classifier is able to decode the auditory emotions and suggests that differences in neural responses to changes in rhythm and timbre between the emotions might contribute to the classifier's performance.

Table 2
Classification of emotion of stimuli using acoustic features (duration and fundamental frequency).

		Happy	Sad	Fear	Total
Within classification	w/duration	0.73	0.83	0.90	0.82
	w/out duration	0.53	0.87	0.77	0.72
Cross-classification:	w/duration	0.50	0.20	0.90	0.57
Test on voice	w/out duration	0.50	0.40	0.70	0.53
Cross-classification:	w/duration	0.50	0.70	0.30	0.50
Test on violin	w/out duration	0.50	0.30	0.30	0.37
Cross-classification:	w/duration	0.60	1.00	0.60	0.73
Test on clarinet	w/out duration	0.80	1.00	0.60	0.80

Discussion

By using multivariate cross-classification and searchlight analyses with different types of auditory stimuli that convey the same three emotions, we identified higher-level neural regions that process the affective information of sounds produced from various sources. Using fMRI data collected from the entire brain, above-chance classification of emotions expressed through auditory stimuli was found both within and across instruments. Searchlight analyses revealed that the primary and secondary auditory cortices, including the superior temporal gyrus (STG) and sulcus (STS), extending into the parietal operculum and posterior insula, exhibit emotion-specific and modality-general patterns of neural activity. This is supported by the fact that BOLD signal in these regions could differentiate the affective content when the classifier was trained on data from one instrument and tested on data from another instrument. Furthermore, within and cross-modal classification performance within a region spatially confined to the dorsal anterior portion of the insula was positively correlated with a behavior measure of empathy. To our knowledge, this is the first study to report the emotion-related spatial patterns that are shared across both musical instruments and vocal sounds as well as to link the degree of predictive information within these spatial patterns with individual differences.

The findings confirm the role of the cortices in the STG and the STS regions in perceiving emotions conveyed by auditory stimuli. Significant classification of vocal expressions of emotions was previously reported in the STG (Kotz et al., 2013) and the region is active when processing acoustical (Salimpoor et al., 2015) and affective components of music (Koelsch, 2014). The left STG was also found to code for both lower-level acoustic aspects as well as higher-level evaluative judgments of

nonlinguistic vocalizations of emotions (Bestelmeyer et al., 2014). It has been suggested that the STG and STS bilaterally may be involved in tracking the changing acoustic features of sounds as they evolve over time (Schonwiesner et al., 2005). The STS in particular appears to integrate audio and visual information during the processing of non-verbal affective stimuli (Kreifelts et al., 2009). Both facial and vocal expressions of emotions activate the STS (Escoffier et al., 2013; Wegrzyn et al., 2015). Multivariate neuroimaging studies have proposed that supramodal mental representations of emotions lie in the STS (Peelen et al., 2010). Furthermore, aberrations in both white-matter volume (von dem Hagen et al., 2011) and task-based functional activity (Alaerts et al., 2014) in these regions were associated with emotion recognition deficits in individuals with autism spectrum disorder (ASD). Our findings suggest that discrete emotions expressed through music are represented by similar patterns of activity in the auditory cortex as when expressed through the human voice. This confirms the role of the STG and STS in processing the perceived affective content of a range of sounds, both musical and non-musical, that is not purely dependent on lower-level acoustic features.

While the peak of the searchlight accuracy maps was located in the auditory cortex, the significant results extend into the parietal operculum. Because these two regions are adjacent and because the neuroimaging data are spatially smoothed both in the preprocessing steps and in the searchlight analysis, we cannot be certain that the significant classification accuracy found in the parietal operculum indicates that this region is additionally involved in representing emotions from sounds. Nonetheless, the idea that cross-modal representation of emotions could be located in this region is consistent with previous research. The inferior portion of the somatosensory cortex, which is located in the parietal operculum (Eickhoff et al., 2006), has been shown to be engaged during vicarious experiences of perceived emotions (Straube and Miltner, 2011). Furthermore, patients with lesions in the right primary and secondary somatosensory cortices performed poorly in emotion recognition tasks (Adolphs et al., 2000) and reported reduced intensity of subjective feelings in response to music (Johnsen et al., 2009). Transcranial magnetic stimulation applied over the right parietal operculum region was also shown to impede the ability to detect the emotions of spoken language (Rijn et al., 2005). Using multivariate methods, Man et al. (2015) found activity in the parietal operculum could be used to reliably classify objects when presented aurally, visually, and tactilely, suggesting that this region contains modality invariant representations of objects and may therefore serve as a convergence zone for information coming from multiple senses. Taken together, the fact that we find significant predictive affective information in the parietal operculum may suggest that the ability to recognize the emotional content of sounds relies on an internal simulation of the actions and sensations that go into producing such sounds.

The searchlight accuracy maps additionally extended into the posterior portion of the insula. The insula is believed to be involved in mapping bodily state changes associated with particular feeling states (Damasio et al., 2013; Immordino-Yang et al., 2014). A range of subjectively-labeled feeling states could be decoded from brain activity in the insula, suggesting that the physiological experience that distinguishes one emotion from another is linked to distinct spatial patterns of activity in the insula (Saarimaki et al., 2015). Studies have shown that the region is largely modality invariant, activated in response to facial expressions of emotions (Wegrzyn et al., 2015), perceptual differences between emotions conveyed through non-speech vocalizations (Bestelmeyer et al., 2014), multimodal presentations of emotions (Schirmer and Adolphs, 2017) and by a wide range of emotions conveyed through music (Baumgartner et al., 2006; Park et al., 2013). The insula's role in auditory processing may be to allocate attentional resources to salient sounds (Bamiou et al., 2003) as evidenced by cases in which patients with lesions that include the insula but not Heschl's gyrus develop auditory agnosia (Fifer et al., 1993). The function of the insula in processing emotions expressed across the senses is further substantiated by the observation that a patient with a lesion in the insula showed an impaired ability to recognize the emotion disgust when expressed in multiple modalities (Calder et al., 2000). Developmental disorders characterized by deficits in emotional awareness and experience may be linked to aberrant functioning of the insula, as decrease insular activity was observed in ASD children observing emotional faces (Dapretto et al., 2006) and altered resting-state functional connectivity between the posterior insula and somatosensory cortices was observed in adults with ASD (Ebisch et al., 2011). The fact that emotions conveyed through auditory stimuli could be classified based on activity in the posterior insula in our study provides further evidence for the hypothesis that perceiving and recognizing an emotion entails recruiting the neural mechanisms that represent the subjective experience of that same emotion.

Despite this finding, classification accuracy within a region of interest in the *dorsal anterior* portion of the insula was significantly positively correlated with empathy. The anterior insula was not one of the significant regions found in the searchlight analysis. These two results might be explained in the context of previous functional and structural imaging studies that suggest that subdivisions of the insular cortex are associated with specific functions (Deen et al., 2011). According to such accounts, the posterior insula, which is structurally connected to the somatosensory cortices, is more directly involved in affective processing of visceral sensations (Kurth et al., 2010) and interoceptive awareness (Craig, 2009), whereas the dorsal anterior insula, which is connected to the cognitive control network and the ACC, is more directly involved in socio-emotional abilities such as empathy and subjective awareness (Craig, 2009). This is evidenced by the fact that the anterior insula is activated when both observing and imitating the emotions of others (Carr et al., 2003). Measures of empathic concern, a subtype of affective empathy referring to the tendency to feel sympathy or concern for others, have been shown to be positively correlated with anterior insula activity when viewing emotional pictures (Silani et al., 2008) as well as when observing loved-ones in pain (Singer et al., 2004). Given that classification accuracy obtained from data within the dorsal anterior insula specifically, was correlated with empathic concern, our results provide further evidence for the unique role of this subdivision in enabling the emotional resonance that is essential to understanding the feelings of others. We speculate that individuals who readily behave empathically might have more finely tuned representations of emotions in the dorsal anterior insula when processing affective information.

While we were mainly interested in identifying brain regions that conserve affect-related information across sounds with differing acoustical properties, we recognize that certain acoustic properties are also integral to specific emotional categories regardless of the source of the sound. Previous results have shown that happiness, for example, is characterized by higher fundamental frequencies and faster tempos when conveyed through both vocal expressions and musical pieces (Juslin and Laukka, 2003). An earlier attempt to disentangle the neural processing of acoustic changes from the neural processing of perceptual changes associated with two different emotions conveyed through auditory stimuli acknowledged that the two are interrelated and that a complete and straightforward separation of the two would be overly simplistic; indeed, the researchers found evidence for both distinct and overlapping neural networks associated with these two processes (Bestelmeyer et al., 2014). Because of this, we did not intend to control for all potential acoustic differences between our stimuli, believing that these features may be essential to that emotional category. Because the duration of the clips varied significantly by emotion and is a feature not directly tied to affective expression, we did regress out the variance explained by duration from our GLM model and showed that cross-instrument classification performance did not change. Besides for duration, no other acoustic properties of the sounds were regressed out of the signal. We therefore might expect that our classifier may be sensitive to signal that is responsive to certain acoustic variations.

To make predictions about the types of acoustic variation that the classifier may be sensitive to, we conducted classification using several audio features extracted for each clip. We found that the acoustic-based classifier performance was also largely dependent on differences in duration between the emotions, as evidenced by the attenuating in performance when durational information was removed. While it is difficult to know what types of information the fMRI-based classifier is using to make distinctions between the emotional states, the classifier trained on acoustic features alone without duration can provide some hypotheses. Once duration was removed, the most informative features for classification of emotions include a rhythmic feature called fluctuation centroid. These results suggest that the fMRI-based classifier is not only sensitive to BOLD signal corresponding to the duration and frequency of the sounds, and may be capturing finely-tuned responses in the auditory cortex and insula that are sensitive to changes in rhythm and timbre that are integral to conveying emotions through sound.

Using BOLD data from a mask of the entire brain, classification accuracies were between 38 and 43% for the whole-brain within instrument classification and 36–40% for the whole-brain cross-instrument classification. Theoretically, if the classifier was guessing the emotional category at random, just by chance, it would correctly identify the emotion 33% of the time. While we recognize that the accuracies obtained here are not impressively high compared to theoretical chance, they are statistically significant according to a one-way *t*-test corrected for multiple comparisons and comparable to those reported in other multivariate cross-classification fMRI studies (Skerry and Saxe, 2014; Kim et al., 2017). Furthermore, the cross instrument classification accuracies should be interpreted in relation to the within instrument classification accuracies, which may set the upper bound of possible performance of a cross-modal classifier (Kaplan et al., 2015). We therefore would not expect cross instrument classification to perform better than within instrument classification and the fact that the cross instrument accuracies are still significantly above chance provides us with compelling evidence that unique spatial patterns of BOLD signal throughout does contain some predictive information regarding the emotional category.

Of additional note, upon inspection of the confusion matrices, the within-instrument classification performance correctly identified fearful clips to a greater degree than the other two emotions, despite the fact positive predictive value across all three was not drastically different. This contrasts with the behavioral findings, in which fear was the emotion most difficult to identify and label. This could indicate that while the fearful clips were easily distinguishable from the other two emotions, these perceived differences may not necessarily adhere to the concepts and features humans have learned to associate with the categorical label of fear. Further exploration into the acoustic components and behavioral responses to fearful musical and vocal clips will help to interpret these opposing findings.

In sum, our study reveals the emotion-specific neural information that is shared across sounds from musical instruments and the human voice. The results support the idea that the emotional meaning of sounds can be represented by unique spatial patterns of neural activity in sensory and affect processing areas of the brain, representations that do not depend solely on the specific acoustic properties associated with the source instrument. These findings therefore have implications for scientific investigations of neurodevelopmental disorders characterized by an impaired ability to recognize vocal expressions of emotions (Allen et al., 2013) and provide a clearer picture of the remarkable ability of the human brain to instantaneously and reliably infer emotions when conveyed nonverbally.

Funding

Funding for this work was provided by the Brain and Creativity Institute.

Acknowledgements

The authors would like to thank Hanna Damasio for her assistance and input regarding neuroanatomical distinctions and all private donors to the Brain and Creativity Institute.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.02.058.

References

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., Damasio, A.R., 2000. A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. J. Neurosci. 20, 2683–2690.
Alaerts, K., Woolley, D.G., Steyaert, J., Martino, A. Di, Swinnen, S.P., Wenderoth, N., 2014. Underconnectivity of the superior temporal sulcus predicts emotion recognition deficits in autism. Soc. Cog. Affect. Neurosci. 9, 1589–1600.
Allen, R., Davis, R., Hill, E., 2013. The effects of autism and alexithymia on physiological and verbal responsiveness to music. J. Autism Dev. Disord. 43, 432–444.
Alluri, V., Toiviainen, P., Jääskeläinen, I.P., Glerean, E., Sams, M., Brattico, E., 2012. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. Neuroimage 59, 3677–3689.
Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., Armony, J.L., 2013. Fear across the senses: brain responses to music, vocalizations and facial expressions. Soc. Cognit. Affect Neurosci. 10, 399–407.
Bamiou, D., Musiek, F.E., Luxon, L.M., 2003. The insula (Island of Reil) and its role in auditory processing: literature review. Brain Res. Rev. 42, 143–154.
Baumgartner, T., Lutz, K., Schmidt, C.F., Jäncke, L., 2006. The emotional power of music: how music enhances the feeling of affective pictures. Brain Res. 1075, 151–164.
Belin, P., Fillion-Bilodeau, S., Gosselin, F., 2008. The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. Behav. Res. Meth. 40, 531–539.
Bestelmeyer, P.E.G., Maurage, P., Rouger, J., Latinus, M., Belin, P., 2014. Adaptation to vocal expressions reveals multistep perception of auditory emotion. J. Neurosci. 34, 8098–8105.
Calder, A.J., Keane, J., Manes, F., Antoun, N., Young, A.W., 2000. Impaired recognition and experience of disgust following brain injury. Nat. Neurosci. 3, 1077–1078.
Carr, L., Iacoboni, M., Dubeau, M., Mazziotta, J.C., Lenzi, G.L., 2003. Neural mechanisms of empathy in humans : a relay from neural systems for imitation to limbic areas. Proc. Natl. Acad. Sci. 100, 5497–5502.
Craig, A.D., 2009. How do you feel–now? The anterior insula and human awareness. Nat. Rev. Neurosci. 10, 59–70.
Damasio, A., Damasio, H., Tranel, D., 2013. Persistence of feelings and sentience after bilateral damage of the insula. Cerebr. Cortex 23, 833–846.
Dapretto, M., Davies, M.S., Pfeifer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y., Iacoboni, M., 2006. Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. Nat. Neurosci. 9, 28–30.
Davis, M.H., 1983. Measuring individual differences in empathy: evidence for a multidimensional approach. J. Pers. Soc. Psychol. 44, 113–126.
Deen, B., Pitskel, N.B., Pelphrey, K.A., 2011. Three systems of insular functional connectivity identified with cluster analysis. Cerebr. Cortex 21, 1498–1506.
Ebisch, S.J.H., Gallese, V., Willems, R.M., Mantini, D., Groen, W.B., Romani, G.L., Buitelaar, J.K., Bekkering, H., 2011. Altered intrinsic functional connectivity of anterior and posterior insula regions in high-functioning participants with autism spectrum disorder. Hum. Brain Mapp. 32, 1013–1028.
Eickhoff, S.B., Schleicher, A., Zilles, K., 2006. The human parietal operculum. I. Cytoarchitectonic mapping of subdivisions. Cerebr. Cortex 15, 254–267.
Ekman, P., 1992. An argument for basic emotions. Cognit. Emot. 6, 169–200.
Escoffier, N., Zhong, J., Schirmer, A., Qiu, A., 2013. Emotional expressions in voice and music: same code, same effect? Hum. Brain Mapp. 34, 1796–1810.
Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P., 2009. Decoding of emotional information in voice-sensitive cortices. Curr. Biol. 19, 1028–1033.
Fifer, R.C., 1993. Insular stroke causing unilateral auditory processing disorder: case report. J. Am. Acad. Audiol. 4, 364–369.
Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A.D., Koelsch, S., 2009. Universal recognition of three basic emotions in music. Curr. Biol. 19, 573–576.
Frühholz, S., Trost, W., Grandjean, D., 2014. The role of the medial temporal limbic system in processing emotions in voice and music. Prog. Neurobiol. 123, 1–17.
Hailstone, J.C., Omar, R., Henley, S.M.D., Frost, C., Michael, G., Warren, J.D., Hailstone, J.C., Omar, R., Henley, S.M.D., Frost, C., Hailstone, J.C., Omar, R., Henley, S.M.D., Frost, C., Warren, J.D., 2009. It's not what you play, it's how you play it: timbre affects perception of emotion in music. Q. J. Exp. Psychol. 62, 2141–2155.
Heller, R., Stanley, D., Yekutieli, D., Nava, R., Benjamini, Y., 2006. Cluster-based analysis of FMRI data. Neuroimage 33, 599–608.
Immordino-Yang, M.H., Yang, X.-F., Damasio, H., 2014. Correlations between social-emotional feelings and anterior insula activity are independent from visceral states but influenced by culture. Front. Hum. Neurosci. 8, 1–15.
Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.

Johnsen, E.L., Tranel, D., Lutgendorf, S., Adolphs, R., 2009. A neuroanatomical dissociation for emotion induced by music. Int. J. Psychophysiol. 72, 24–33.
Juslin, P.N., Laukka, P., 2003. Communication of emotions in vocal expression and music performance: different channels, same code? Psychol. Bull. 129, 770–814.
Kao, M., Mandal, A., Lazar, N., Stufken, J., 2009. NeuroImage Multi-objective optimal experimental designs for event-related fMRI studies. Neuroimage 44, 849–856.
Kaplan, J.T., Man, K., Greening, S.G., 2015. Multivariate cross-classification : applying machine learning techniques to characterize abstraction in neural representations. Front. Hum. Neurosci. 9, 1–12.
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, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J a, Turnbull, D., 2010. Music Emotion Recognition : a State of the Art Review, pp. 255–266. Inf Retr Boston.
Kleiner, M., Brainard, D.H., Pelli, D., 2007. What's New in Psychtoolbox-3? Percept 36 ECVP Abstr Suppl.
Koelsch, S., 2014. Brain correlates of music-evoked emotions. Nat. Rev. Neurosci. 15, 170–180.
Kotz, S.A., Kalberlah, C., Bahlmann, J., Friederici, A.D., Haynes, J.D., 2013. Predicting vocal emotion expressions from the human brain. Hum. Brain Mapp. 34, 1971–1981.
Kragel, P.A., LaBar, K.S., 2015. Multivariate neural biomarkers of emotional states are categorically distinct. Soc. Cognit. Affect Neurosci. 10, 1437–1448.
Kreifelts, B., Ethofer, T., Grodd, W., 2009. Cerebral representation of non-verbal emotional perception : fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. Neuropsychologia 47, 3059–3066.
Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based Functional Brain Mapping.
Kurth, F., Zilles, K., Fox, P.T., Laird, A.R., Eickhoff, S.B., 2010. A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. Brain Struct. Funct. 214, 519–534.
Lamm, C., Batson, C.D., Decety, J., 2007. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. J. Cognit. Neurosci. 19, 42–58.
Lartillot, O., Toiviainen, P., 2007. A Matlab toolbox for musical feature extraction from audio. In: International Conference on Digital Audio Effects, pp. 237–244.
Linke, A.C., Cusack, R., 2015. Flexible information coding in human auditory cortex during perception, imagery, and STM of complex sounds. J. Cognit. Neurosci. 27.
Man, K., Damasio, A., Meyer, K., Kaplan, J.T., 2015. Convergent and invariant object representations for Sight,Sound, and touch. Hum. Brain Mapp. 36, 3629–3640.
Mullensiefen, D., Gingas, B., Musil, J., Steward, L., 2014. The musicality of non-musicians: an Index for assessing musical sophistication in the general population. PLoS One 9.
Norman, K a, Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cognit. Sci. 10, 424–430.
Paquette, S., Peretz, I., Belin, P., 2013. The " Musical Emotional Bursts ": a validated set of musical affect bursts to investigate auditory affective processing. Front. Psychol. 4, 1–7.
Park, M., Hennig-Fast, K., Bao, Y., Carl, P., Pöppel, E., Welker, L., Reiser, M., Meindl, T., Gutyrchik, E., 2013. Personality traits modulate neural responses to emotions expressed in music. Brain Res. 1523, 68–76.
Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal representations of perceived emotions in the human brain. J. Neurosci. 30, 10127–10134.
Rigoulot, S., Pell, M.D., Armony, J.L., 2015. Time course of the influence of musical expertise on the processing of vocal and musical sounds. Neuroscience 290, 175–184.
Rijn, S Van, Aleman, A., Diessen, E Van, Berckmoes, C., Vingerhoets, G., Kahn, R.S., 2005. What is said or how it is said makes a difference : role of the right fronto-parietal operculum in emotional prosody as revealed by repetitive TMS. Eur. J. Neurosci. 21, 3195–3200.
Saarimaki, H., Gotsopoulos, A., Jaaskelainen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., Nummenmaa, L., 2015. Discrete neural signatures of basic emotions. Cerebr. Cortex 1–11.
Salimpoor, V.N., Zald, D.H., Zatorre, R.J., Dagher, A., Mcintosh, A.R., 2015. Predictions and the brain : how musical sounds become rewarding. Trends Cognit. Sci. 19, 86–91.
Sander, K., Scheich, H., 2005. Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. J. Cognit. Neurosci. 17, 1519–1531.
Schirmer, A., Adolphs, R., 2017. Emotion perception from Face,Voice, and touch: comparisons and convergence. Trends Cognit. Sci. 21, 216–228.
Schonwiesner, M., Rübsamen, R., von Cramon, D.Y., 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur. J. Neurosci. 22, 1521–1528.
Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., Frith, U., 2008. Levels of emotional awareness and autism : an fMRI study. Soc. Neurosci. 3, 97–112.
Singer, T., Seymour, B., Doherty, J.O., Kaube, H., Dolan, R.J., Frith, C.D., 2004. Empathy for pain involves the affective but not sensory components of pain. Science (80-) 303, 1157–1162.
Skerry, A.E., Saxe, R., 2014. A common neural code for perceived and inferred emotion. J. Neurosci. 34, 15997–16008.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208–S219.
Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44, 83–98.
Stelzer, J., Chen, Y., Turner, R., 2013. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage 65, 69–82.
Straube, T., Miltner, W.H.R., 2011. Attention to aversive emotion and specific activation of the right insula and right somatosensory cortex. Neuroimage 54, 2534–2538.
von dem Hagen, E.A.H., Nummenmaa, L., Yu, R., Engell, A.D., Ewbank, M.P., Calder, A.J., 2011. Autism spectrum traits in the typical population predict structure and function in the posterior superior temporal sulcus. Cerebr. Cortex 21, 492–500.
Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., Bien, C.G., Kissler, J., 2015. Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. Cortex 69, 131–140.
Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M., Nichols, T.E., 2014. Permutation inference for the general linear model. Neuroimage 92, 381–397.