Brain and Cognition 101 (2015) 1–11

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Music-induced emotions can be predicted from a combination of brain activity and acoustic features

Ian Daly [a,*], Duncan Williams [b], James Hallowell [a], Faustina Hwang [a], Alexis Kirke [b], Asad Malik [a], James Weaver [a], Eduardo Miranda [b], Slawomir J. Nasuto [a]

[a] Brain Embodiment Lab, School of Systems Engineering, University of Reading, Reading, UK
[b] Interdisciplinary Centre for Music Research, University of Plymouth, Plymouth, UK

ARTICLE INFO

Article history:
Received 23 January 2015
Revised 3 August 2015
Accepted 4 August 2015
Available online 3 November 2015

Keywords:
Music
Affective state prediction
EEG
Acoustic features
Machine learning

ABSTRACT

It is widely acknowledged that music can communicate and induce a wide range of emotions in the listener. However, music is a highly-complex audio signal composed of a wide range of complex time- and frequency-varying components. Additionally, music-induced emotions are known to differ greatly between listeners. Therefore, it is not immediately clear what emotions will be induced in a given individual by a piece of music.

We attempt to predict the music-induced emotional response in a listener by measuring the activity in the listeners electroencephalogram (EEG). We combine these measures with acoustic descriptors of the music, an approach that allows us to consider music as a complex set of time-varying acoustic features, independently of any specific music theory. Regression models are found which allow us to predict the music-induced emotions of our participants with a correlation between the actual and predicted responses of up to $r = 0.234, p < 0.001$.

This regression fit suggests that over 20% of the variance of the participant's music induced emotions can be predicted by their neural activity and the properties of the music. Given the large amount of noise, non-stationarity, and non-linearity in both EEG and music, this is an encouraging result. Additionally, the combination of measures of brain activity and acoustic features describing the music played to our participants allows us to predict music-induced emotions with significantly higher accuracies than either feature type alone ($p < 0.01$).

1. Introduction

Music is widely acknowledged to be a powerful method for emotional communication, capable of eliciting a range of different emotional responses in the listener, such as joy, excitement, and fear (Scherer, 2004). Subsequently, music therapy may be used as a tool for treatment of emotional disorders such as depression (Maratos, Gold, Wang, & Crawford, 2008).

Music therapy is a health intervention in which the music therapist uses music as a tool to help their patient with their physical and/or mental health problems (Bradt, Magee, Dileo, Wheeler, & McGilloway, 2010; Erkkilä et al., 2011; McDermott, Crellin, Ridder, & Orrell, 2013). For example, in the treatment of depression music therapy has been reported to significantly improve mood when compared to standard care alone (Maratos et al., 2008) (for example antidepressant drugs alone vs. antidepressant drugs and music therapy (Chen, 1992)).

The music used in music therapy is selected by the therapist based upon a combination of the therapists evaluation of their patients current psychological state, the therapists expertise and experience, and the properties of the music that the therapist judges will be beneficial to the patient (Tamplin & Baker, 2006).

In order to select an appropriate piece of music for use in music therapy it is necessary to predict how the individual is likely to react to that piece of music. However, it is a considerable challenge to predict the potential reaction of an individual to a piece of music they have not previously heard before. There are large interpersonal differences in emotions induced by listening to a piece of music, which result from both the music itself and the participant's own previous and current mental states (Hunter, Schellenberg, & Schimmack, 2010).

These inter-person differences are a result of a wide range of influences and include the individuals prior experiences, their current mood, and a range of other factors both internal to the person

* Corresponding author.
E-mail address: i.daly@reading.ac.uk (I. Daly).

http://dx.doi.org/10.1016/j.bandc.2015.08.003
0278-2626/

and external to them. Broadly speaking, a persons emotional response to a piece of music can be said to be a function of both the music itself and of the individual.

When considering the piece of music, a number of models have been proposed for the relationships between musical structure and syntax and both the perceived and/or induced emotional responses of a listener (for example, Thompson & Robitaille (1992), Schubert (1999), Gabrielsson & Lindström (2001), Gabrielsson & Juslin (2003)).

For example, in Russell (1980) and Livingstone and Thompson (2009) the circumplex model of affect and its relationship to musical descriptors is described. In this model, emotional responses have been plotted across two continuous axes, arousal (excitement) and valence (pleasantness), ranging from low to high. Musical descriptors drawn from music theory, such as tempo or modality, are plotted in this space.

However, while this model is intuitive and can be informative about perceptions of the role individual features of music theory may play in emotional responses, it is not complete. First, due to the very large inter-person differences in music-induced emotions, music features are not likely, by themselves, to be good predictors of emotional responses to music. This may be due to a variety of factors, including individual preferences for particular pieces of music, prior experience of music induced emotions, or a participants physiological state as they listen to music (Peretz, Aube, & Armony, 2013).

An alternative approach that has been adopted is to attempt to use physiological measurements of the participant as correlates of their emotional responses (Craig, 2005; Kim & André, 2008; Schmidt & Trainor, 2001). Patterns in these physiological measurements can be identified and used to attempt to identify a participants emotional response to a piece of music.

An example of this is the use of electrocardiogram (ECG) signals to identify emotional responses to music (Kim & André, 2008). As music causes listeners to become more excited, this can lead to increases in heart rate, which is reflected in the recorded ECG signal and, subsequently, classified (Kim & André, 2008). Other physiological measures which have been adopted to identify music-induced emotions include the galvanic skin response (GSR) (Craig, 2005), the electromyogram (EMG) recorded from the facial muscles (Lundqvist, Carlsson, Hilmersson, & Juslin, 2008), and respiration rate (Etzel, Johnsen, Dickerson, Tranel, & Adolphs, 2006).

Alternatively, a number of researchers have explored various indices of neural activity as a measure of music-induced emotion. This may be done by, for example, the use of the electroencephalogram (EEG) (Schmidt & Trainor, 2001).

Measures of activity in the EEG which have been reported to relate to music-induced emotion include asymmetry of activity within the alpha band over the prefrontal cortex (Schmidt & Trainor, 2001), measures of prefrontal asymmetry in the beta frequency band, and measures of connectivity between prefrontal and occipital cortical areas (Daly et al., 2014).

The influence of music-induced emotion on the EEG is derived from the neurobiological mechanisms mediating interactions of music with emotions (Peretz, 2009, chap. 5). Music is thought to engage a diverse network of neural structures, with no single pathway bearing responsibility for music-induced emotions. This is evidenced by the lack of reports of selective loss of all music-induced emotions due to brain injury, contrasting with the prevalence of evidence for selective loss of some music-induced emotions. For example, 'scary' and 'sad' music-induced emotions may be lost after damage to the amygdala (Gosselin, 2005; Gosselin, Peretz, Johnsen, & Adolphs, 2007) and impaired by Parkinson's disease (van Tricht, Smeding, Speelman, & Schmand, 2010). This is also evidenced by findings that preferred musical styles engage a listener's default mode network most strongly (Wilkins, Hodges, Laurienti, Steen, & Burdette, 2014).

As a consequence, music-induced emotions relate to a range of particular effects in the EEG. These include inter-hemispheric differences in EEG activity levels (Daly et al., 2014; Flores-Gutiérrez et al., 2007; Schmidt & Trainor, 2001) or changes in EEG over specific regions, such as the pre-frontal cortex (Lin et al., 2010). Taken together, it has been suggested that musical emotions engage a network of both cortical and sub-cortical regions, which produces a range of effects in the EEG (Peretz, 2009, chap. 5).

These effects are widely known to differ between individuals (Hunter et al., 2010). This can occur for a variety of reasons, including musical preferences (Bauer, Kreutz, & Herrmann, 2015), age (Daly et al., 2014), and emotional intelligence (Jausovec & Jausovec, 2005). Additionally, the EEG is known to be a noisy, non-stationary signal (Daly et al., 2012). Taken together this makes reliable identification of music-induced emotions from the EEG a very challenging problem.

Therefore, we suggest that a combination of both physiological measures of the listener and acoustic properties of the music may be used to effectively predict emotional responses to a piece of music. Specifically, we hypothesise that a combination of EEG measures and the acoustic properties of the music may be used to predict the emotional response they will report experiencing while listening to the music.

We play a series of musical clips to a group of participants, while recording their EEG. We then extract descriptive features of both the acoustic properties of the music and the participant's EEG. We attempt to use these features to train a regression model to predict the music-induced emotional responses of the participants.

2. Methods

2.1. Measurements

Thirty-one individuals between the ages of 18–66 (median 35, 18 female) participated in the study (previously detailed in (Daly et al., 2014)). All participants were healthy adults who did not report having any mental health, mood, or psychiatric problems. All participants had normal, or corrected to normal, hearing and vision. Twenty-nine of the participants were right handed (no significant differences were found in the results from the two left handed participants). The electroencephalogram (EEG) was recorded from each participant from 19 channels positioned according to the International 10/20 system for electrode placement.

The participants each listened to 40 pieces of music, which were uniformly drawn from a set of 110 excerpts from film scores. The stimuli were taken from a dataset of musical pieces chosen with the specific purpose of inducing emotional responses in the listener (Eerola & Vuoskoski, 2010).

Each musical clip was played for 12 s, as described in Daly et al. (2014), during which the participants were instructed to look at the screen and listen to the music without moving. They were then asked a series of 8 randomly-ordered Likert questions designed to identify the level of emotional response along 8 axes induced in them by the music.

These 8 axes allowed the participants to report their music-induced emotions in terms of pleasantness, energy, sadness, anger, tenderness, happiness, fear, and tension. However, as several of these categories are likely to be highly correlated, a principal component analysis (PCA) was used in order to identify a reduced set of categories. Three principal components (PCs) were identified, which explained >75% of the variance of the participant's responses. These three PCs are used in subsequent analysis and referred to as the 'response PCs'. They correspond to each of the axes of the three-dimensional Schimmack and Grob model of affective states (valence, energy-arousal, and tension-arousal) (Schimmack & Grob, 2000). Further details on the measurement procedure and the experimental paradigm are reported in Daly et al. (2014).

From the recorded dataset we extract acoustic features from each of the pieces of music played to the participants. We also extract physiological features from the participant's EEG during each music listening trial. We then attempt to identify subsets of these features which can be used to reliably predict a participant's reported emotional response to the music along each of the chosen response PCs.

2.2. Acoustic features

Each of the musical clips used as stimuli may be described by a range of acoustic features. We select a subset of acoustic features based upon the taxonomy of musical features described in Mitrović, Zeppelzauer, and Breiteneder (2010) and designed to cover the following key musical properties and acoustic feature types: temporal features, spectral features, perceptual features, cepstral features, and features describing the beat of the music.

The acoustic features are extracted from each of the musical stimuli using Matlab and toolboxes Mitrović et al. (2010) and Dubnov (2006). In total 135 acoustic features were extracted from the music.

2.2.1. Temporal features

Temporal features refer to time-varying characteristics extracted from a signal. Temporal features may describe aspects of the amplitude and/or the energy of the signal. In this study we use the following summary temporal features: zero crossings (Kedem, 1986), an amplitude descriptor (Mitrovic, Zeppelzauer, & Breiteneder, 2006), the short-time energy (Zhang & Kuo, 2001), and beats per minute.

Zero crossings rate (ZCR) is defined as the number of zero crossings within the audio signal time series in a fixed time window of length W, which is slid over the length of the signal with no overlap. The ZCR has been described as a measure of the dominant frequency of the signal (Kedem, 1986). It has been used as a feature in a range of problems, for example in music genre classification (Martin Mckinney, 2003). However, for complex music waveforms, it is unclear whether ZCR alone will provide an adequate description of the music such that the emotional response of a participant can be predicted.

Amplitude descriptors separate the audio signal into segments of low and high absolute amplitude via adaptive thresholding, in which the threshold is adapted based upon the current mean amplitude of the signal. The descriptor is then composed of the duration, variation in duration, and independent energy of these segments. This provides a description of the sound in terms of its envelope. Amplitude descriptors have been implemented in animal sound recognition but can also be used for describing audio signals such as music. This suggests that they are readily adaptable to complex waveform analysis such as music (Mitrović et al., 2010).

Short-time energy provides a description of the signal envelope. As recommended in Mitrović et al. (2010), we used the definition from Zhang and Kuo (2001) in which short-time energy is defined as the mean energy per window of length W, which is slid over the signal with no overlap.

Beats per minute may be used to describe the tempo of music. Beats represent a measure of tempo of the audio signal that is a way of measuring the change in the patterns of energy in the music over time (Pampalk, Rauber, & Merkl, 2002).

Beat tracking of the music clips is performed via the dynamic programming approach proposed in Ellis (2007). The mean and standard deviation of the beats per minute are estimated from each music clip over the entire duration of the musical stimuli.

2.2.2. Spectral features

Frequency-based features describe a signal in terms of its spectral content. Thus, the signal must be first translated into the frequency domain, for example via application of a Fourier or wavelet transform. Descriptions of the spectral content of the signal can then be extracted. Physical frequency features refer to the physical properties of the signal, as opposed to how listeners may perceive the signal (Mitrović et al., 2010). The spectral features of interest are spectral centroid (with the semantic meaning of brightness) (Scheirer & Slaney, 1997), autoregressive features (Rabiner, 1979), Daubechies wavelet coefficient histogram (Li, Ogihara, & Li, 2003), spectral flux (Scheirer & Slaney, 1997), spectral slope (Morchen, Ultsch, Thies, & Lohken, 2006), and cepstral features, specifically the Mel-cepstral coefficients.

The spectral centroid is defined as the centre of gravity of the magnitude spectrum and is used to identify the point in the frequency spectrum of the signal with the greatest concentration of energy. Spectral centroid provides a measure of brightness of the signal, where brightness describes whether an audio signal is dominated by low or high frequencies. The greater the dominance of high frequencies in the signal the higher the brightness (Scheirer & Slaney, 1997).

Autoregressive features attempt to describe an audio signal by how well a linear predictor may estimate each value in the signal based upon previous values. Thus, this provides a measure of predictability and stability in the signal over time (Mitrović et al., 2010).

Daubechies wavelet coefficient histogram features (DWCH) provide a measure of the mean frequency content of the audio signal in a set of discrete frequency bands. From each sub-band, the first three statistical moments describe the energy and variation and comprise a measure of the energy per sub-band over time. DWCH features have been used in a number of applications, including genre classification (Li et al., 2003).

Spectral flux is defined as the Euclidean norm of the window-to-window differences in spectral amplitude. Spectral flux may be used to measure the rate of change of the spectrum of the signal over time. Audio signals with lots of large changes in spectrum will have a high spectral flux, while audio with only a small amount of change will have a low spectral flux (Scheirer & Slaney, 1997). Spectral flux is used in a number of applications, including retrieval of musical structure (Li & Ogihara, 2004).

Spectral slope attempts to approximate the shape of the spectrum by applying a linear regression. The angle of the slope represents the change in frequency content of the signal from low to high frequencies and may be used as an alternative feature to identify the relative frequency content of the signal (Morchen et al., 2006).

Cepstral features are defined as frequency-smoothed representations of the log magnitude spectrum that aim to capture timbral characteristics and pitch of the signal (Davis & Mermelstein, 1980). They are widely used in speech, music, and environmental noise processing. In this work we employ Mel-frequency cepstral coefficients (MFCCs) as features to describe the audio signals (Stevens, 1937).

MFCCs are computed by first converting Fourier coefficients of the signal to Mel-scale, where a Mel refers to a difference in pitch that is noticeable to a human listener. The resulting vectors are then logarithmized and decorrelated to remove redundant information. The MFCCs capture the timbre and pitch of the signal by providing a representation of the shape of the spectrum (Stevens, 1937).

2.2.3. Spectro-temporal features

The set of perceptual frequency features used in this analysis contains features for which there is a specific semantic meaning that may be attached to the feature. Thus, these features are relevant to the human auditory perception of sound. The spectral roll off point (Scheirer & Slaney, 1997), specific loudness sensation (Pampalk et al., 2002), and Shepard (1964) are investigated.

Spectral roll off is defined as the frequency below which 95% of the content of the power spectrum is located. Spectral roll off provides a measure of tonality of the signal (Mitrović et al., 2010). Tonality may be described as an attempt to differentiate tonal sounds from noise-like sounds and may be measured by looking at the flatness of the spectrum. The flatter the spectrum the more noise-like. Spectral roll off as a tonality measure has been used in music information retrieval (Morchen et al., 2006).

Specific loudness sensation measures the perceived loudness of an audio signal. This is done by first computing a Bark-scaled spectrogram before applying spectral masking to extract a measure of loudness sensation (Pampalk et al., 2002). The bands of the spectrogram reflect characteristics of the cochlea and inner ear of the auditory system, while the spectral masking reflects the occlusion of quiet sounds by louder sounds when both are present at similar frequencies (Fastl & Zwicker, 2007).

Chroma is used to measure the pitch of an audio signal. This is done by measuring a chromagram from the signal, a measure of the spectral energy of the signal at each one of 12 different pitch classes. This measure is based upon the short time Fourier transform (Goto, 2006). Chroma also provides a general description of the music content at different pitches. Therefore, it is independent of a particular musical theory.

2.3. EEG features

2.3.1. Pre-processing

Prior to use of the EEG signals for analysis we first attempted to remove artefacts from the signals. This was done by first visually inspecting the EEG and manually labelling portions of the data that contained artefacts. Independent component analysis (ICA) was then used to separate the EEG into components containing EEG data and components containing artefacts.

Artefact-contaminated components were then identified via visual inspection and removed before reconstruction of the cleaned EEG. Trials within the data were then marked for inclusion in the analysis if they were not previously labelled as containing electromyogram (EMG) artefacts and they did not contain any amplitude values greater than ± 100 uV.

This resulted in a total of 31.03% of the trials been removed and left 800 artefact-free trials for analysis. Further details of this artefact removal process are reported in Daly et al. (2014).

2.3.2. Feature extraction

Two different types of features were extracted from the EEG; band-power features and pre-frontal asymmetry features.

Band power features were extracted from each of the 19 channels by taking the mean band power from 0–12 s relative to the start time of the music. Twenty non-overlapping frequency bands of width 4 Hz were used from 0 Hz to 80 Hz. Pre-frontal asymmetry is defined here as the difference between the EEG band-power activity on channel F3 and the EEG band-power activity on F4 within each frequency band.

This results in a set of 400 unique features describing the EEG activity during each trial.

2.4. Feature search

The set of acoustic features and EEG features are combined to make a set of 535 candidate features (400 EEG features and 135 acoustic features). We attempt to identify a subset of these features for use in predicting participant-reported music-induced emotions.

To do this a feature selection method based upon principal component analysis is used (Daly et al., 2014). This method first uniformly re-distributes the candidate feature set and coarse grains it. The covariance matrix is then found between all candidate features and an additional vector, appended to the candidate feature set, containing the responses PC currently of interest.

Principal component analysis (PCA) is applied to identify the direction of maximum variance. A participation index is then calculated, which defines the involvement of each principal component with the vector containing the response PC. The top 5'th percentile of these participation indices identify the feature set which is most suitable for identifying the corresponding emotional responses to the music reported by the participants.

Further details on the method are reported in Daly et al. (2014).

2.5. Prediction

A model is sought that predicts the emotional responses of the participants to the music in terms of each of the response PCs. Linear regression models are fitted to the emotional content of the music, as reported by the participants and recorded in the response PCs.

For each response PC a linear regression model is sought that maximises the amount of variance in the response PCs explained by the selected features. This is used to suggest which features will have the greatest impact on the emotional responses reported by the listeners.

2.6. Model training

A 10×10 cross-fold train and validation scheme is used to first identify subsets of features which best relate to each of the response PCs and, second, train a regression model using these response PCs. Thus, separate cross-fold procedures are run to find linear regression models that fit to each of the three response PCs.

Within the training set, in each fold, the regression model is trained by a stepwise training process, which iteratively considers combinations of the selected features as terms for use in the model. After training, the model is applied to attempt to predict the response PCs to each of the trials within the held-out testing set.

2.7. Evaluating the results

The performance of the prediction method is evaluated by identifying how close the predicted response PCs are to the actual recorded response PCs. This is done for each item in the testing set within each fold of the cross-fold train and validation scheme.

Performance is evaluated by calculating the correlation between the actual response PCs and the predicted response PCs. The statistical significance of this correlation is estimated parametrically.

3. Results

The performance of the prediction approach is evaluated first for the combination of both EEG features and acoustic features. It is then evaluated for EEG features and acoustic features separately.

3.1. Combined features

Each response PC is considered individually. Feature subsets are sought that can be used to predict the responses of participants to the music in each trial.

For each of the emotional response PCs the mean correlation between the predicted response PC and the actual response PC is listed in Table 1. In each case the selected feature set and trained regression model are able to predict the response PCs with highly significant correlations ($p < 0.01$). The r-values of the correlation are relatively high, given the widely-reported high levels of noise and non-stationarity in the EEG (Schomer, Blum, & Rutkove, 2007).

Additionally, it is noted that the participants have a wide range of ages. Therefore, it is important to consider whether participant age significantly affects the results. To this end the prediction was also attempted on a per participant basis and correlations were calculated between prediction results and participant ages for each of the three emotional response PCs.

After correcting for multiple comparisons (Bonferroni correction, $N = 3$ for the 3 emotional response PCs), no significant correlations were found between prediction performance and the participant's ages ($p > 0.05$). This confirms that age does not significantly affect the results.

The number of artefact contaminated trials removed from the dataset may also influence our results. To investigate this, correlations were calculated between the number of artefacts removed from the EEG recorded from each participant and the prediction results for each emotional response PC. No significant correlations were found ($p > 0.05$), suggesting that the number of artefact contaminated trials removed from the EEG does not affect the results.

Fig. 1 illustrates the correlations between the predicted and actual emotional responses to music reported by the participants when a combination of both EEG and acoustic features are used to train the regression models. Note that for each response PC the correlation between the predicted and the actual response PC is positive.

The features identified for use in predicting the emotional responses to the music are illustrated in Fig. 2. Each feature is labelled and the percentage of folds in which it is selected is indicated by its height on the y-axis. The features are included in the figure if they are selected in more than 5% of the folds of the cross-fold train and validation scheme.

The EEG features which are selected are then investigated. Fig. 3 illustrates from which frequency bands the mean band-power features were selected in order to predict the response PCs. Note that response PC1 and response PC3 are predicted best by activity in the delta and beta frequency bands, while response PC2 is predicted best by activity in the beta and gamma bands.

Fig. 4 illustrates the scalp maps of the spatial locations from which these features are selected for each of the response PCs. From these maps it may be observed that response PC1 is predicted by band-power features measured over the prefrontal cortex and right-frontal cortex, while PC2 is predicted best by band-power features measured over the left motor cortex, and response PC3 is predicted best by band-power features measured over the right motor cortex.

Prefrontal asymmetry is also investigated as a potential feature and is selected by our feature selection method for use in the regression models. Fig. 5 illustrates a histogram of the band-powers from which asymmetry features were selected for use in predicting the response PCs. Asymmetry in the beta band may be observed to be a good predictor of response PC1, while asymmetry in the alpha band is a good predictor of response PC3.

Additionally, for each response PC several acoustic features are also selected for use in prediction. These are illustrated in Fig. 6. High frequency Mel-cepstral coefficients are observed to be a good predictor of response PC1, while low frequency Mel-cepstral coefficients and Chroma features are occasionally selected to predict response PC2 and response PC3.

The effects of each of the feature types are now evaluated. The training and testing process is repeated with (a) just features extracted from the EEG and, (b) just acoustic features. For each of these feature types, and for the combined feature set, the training and testing cross-validation scheme is repeated 100 times with different, randomly generated, folds in each iteration. The distribution of the resulting prediction performance (as measured by the correlation between the predicted and actual response PCs) is used to determine whether the prediction performance differs significantly between the three different feature sets.

A 1×3 ANOVA is applied with factor 'features' and levels 'EEG', 'acoustic', and 'combined'. A significant effect of 'features' is found for response PC1 ($F(2,297) = 53250.4, p < 0.0001$), response PC2 ($F(2,297) = 28040.6, p < 0.0001$), and response PC3 ($F(2,297) = 1812.51, p < 0.0001$). In all cases post hoc t-tests reveal significant differences between all groups ($p < 0.01$). Therefore, a combination of EEG and acoustic features results in significantly better prediction of music-induced emotion than either feature type alone.

Table 1
Mean and standard deviation of the performance of the regression models at predicting the response PCs when the models are trained on a combination of EEG and acoustic features, EEG features alone, or acoustic features alone. All results are highly statistically significant ($p < 0.001$).

Response PC	Mean correlation (± std.)		
	Combined	EEG	Acoustic
PC1 (Valence)	0.243 (0.005)	0.202 (0.005)	0.028 (0.003)
PC2 (Energy-arousal)	0.158 (0.006)	0.147 (0.004)	0.018 (0.005)
PC3 (Tension-arousal)	0.102 (0.005)	0.088 (0.022)	0.001 (0.004)

4. Discussion

An individuals emotional response to music is a result of a complex series of interlocking factors. In addition to the acoustic properties of the music being played to the individual, other factors such as the individual's mood, memories, and their current level of engagement may all effect how they respond to hearing a piece of music (Hunter et al., 2010).

By using both features derived from the electroencephalogram (EEG) and acoustic features derived from the music itself we are able to predict a participants emotional response to music significantly more accurately than using either of those feature types alone. This suggests that emotional responses to listening to music are the result of processes that are both internal to the listener and a result of the acoustic properties of the music, i.e. the stimuli presented to the listener.

Subsets of features have been found which allow us to train linear regression models to predict participants responses along each of the first 3 response PCs we have identified. These response PCs correspond to each of the axes of the three dimensional Schimmack and Grob model of affective states (valence, energy-arousal, and tension-arousal) (Schimmack & Grob, 2000; Daly et al., 2014). Thus, our results indicate that we can predict our participants responses along each of these axes.

The valence (response PC1) reported by the participants is observed to be related to features related to the variance of Mel cepstral coefficients in high frequency bands. Energy-arousal is observed to be weakly related to the variance of low frequency Mel cepstral coefficients. Additionally, tension-arousal is observed to be weakly related to Chroma in low frequency bands.

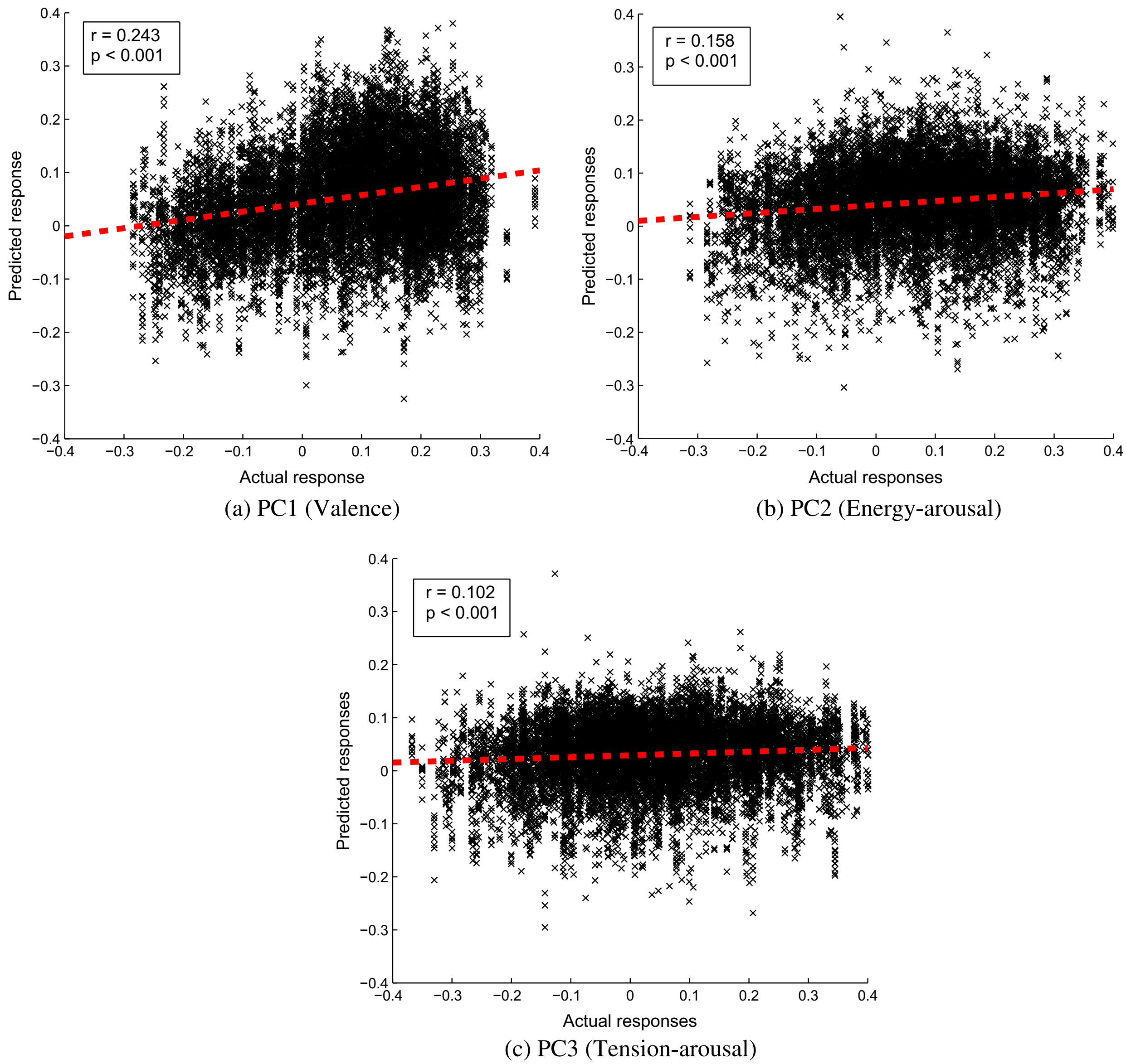


Fig. 1. Correlation between predicted and actual emotional responses to music when the regression model is trained with a combination of EEG features and acoustic features. Blue bars indicate EEG features, while red bars indicate acoustic features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

High frequency Mel-cepstral coefficients correspond to higher pitch instruments and the relationship we identify between this and valence suggests that variance in higher pitches in music predicts valence reported by the listener. Variance of low pitch class Chroma relates to lower pitch keys and our results suggest that more changes in these keys induce changes in tension in the listener.

Valence (response PC1) is observed to be best predicted by EEG features in the delta (0–4 Hz) and beta (13–30 Hz) frequency bands. It is also observed to be predicted by band-powers recorded over the right frontal cortex. Additionally, prefrontal asymmetry measures in the beta frequency band are observed to predict valence (response PC1).

This broadly matches results reported elsewhere. For example, in Walpulski (2008) delta, alpha, and beta frequency bands were observed to most strongly correlate with visually-induced emotional responses, while in Daly et al. (2014) beta frequency bands were observed to correlate with music-induced emotional responses.

Additionally, it may be noted that the observed greater involvement of the right hemisphere in processing music-induced changes in valence reflects results reported elsewhere in the literature. Specifically, the right hemisphere is reported to be most involved in emotional processing in Flores-Gutiérrez et al. (2007) and Silberman and Weingartner (1986). Music-induced emotion is also reported to produce asymmetry effects over the prefrontal cortex, reflecting a form of hemispheric specialisation (Schmidt & Trainor, 2001).

Energy-arousal (response PC2) is observed to be predicted by beta and gamma band-powers. These are observed to be concentrated over the centre of the frontal cortex and over the left motor cortex. Additionally, prefrontal asymmetry is observed to be a very poor predictor of energy-arousal and is selected in only 10% of the folds.

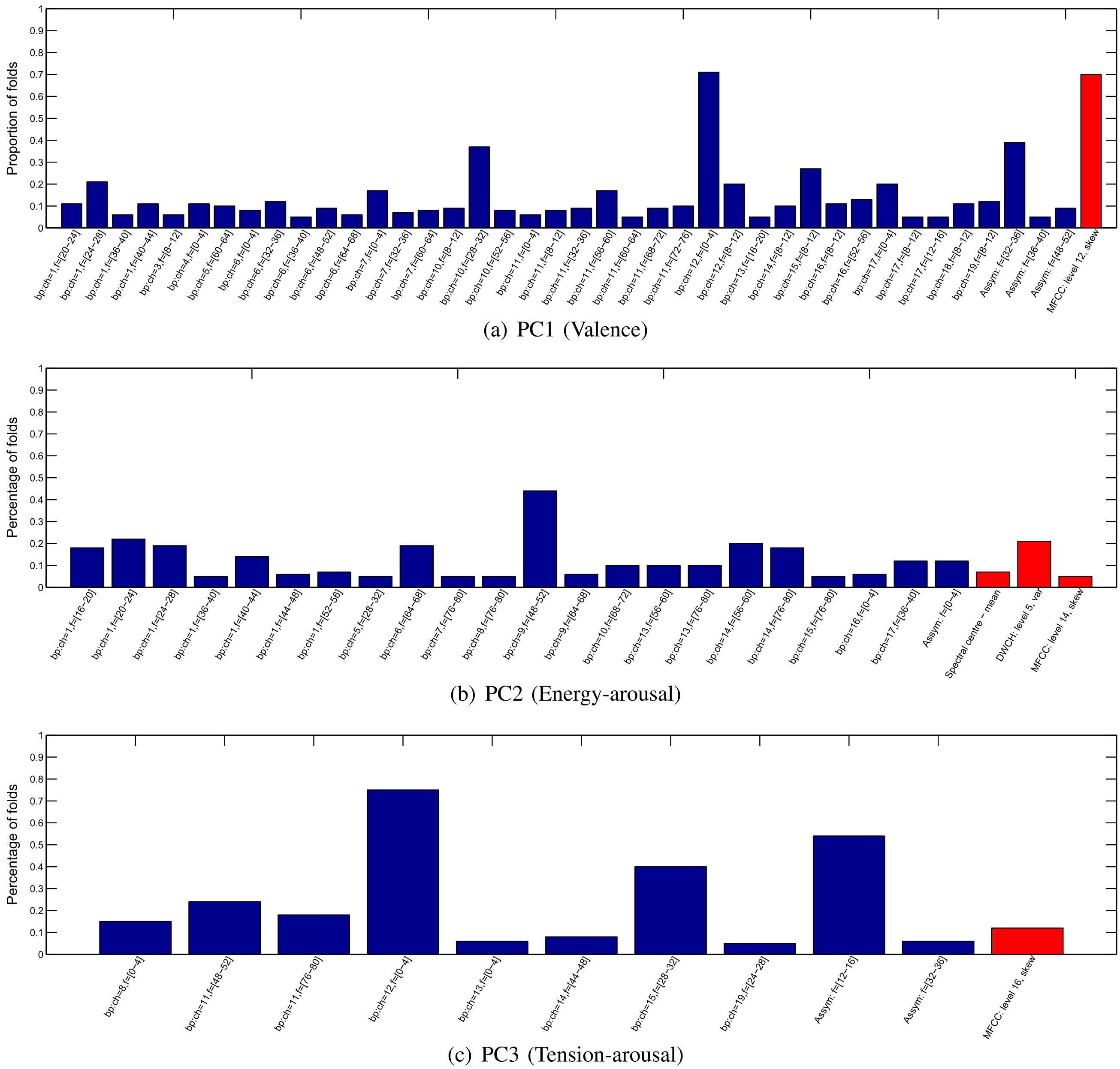


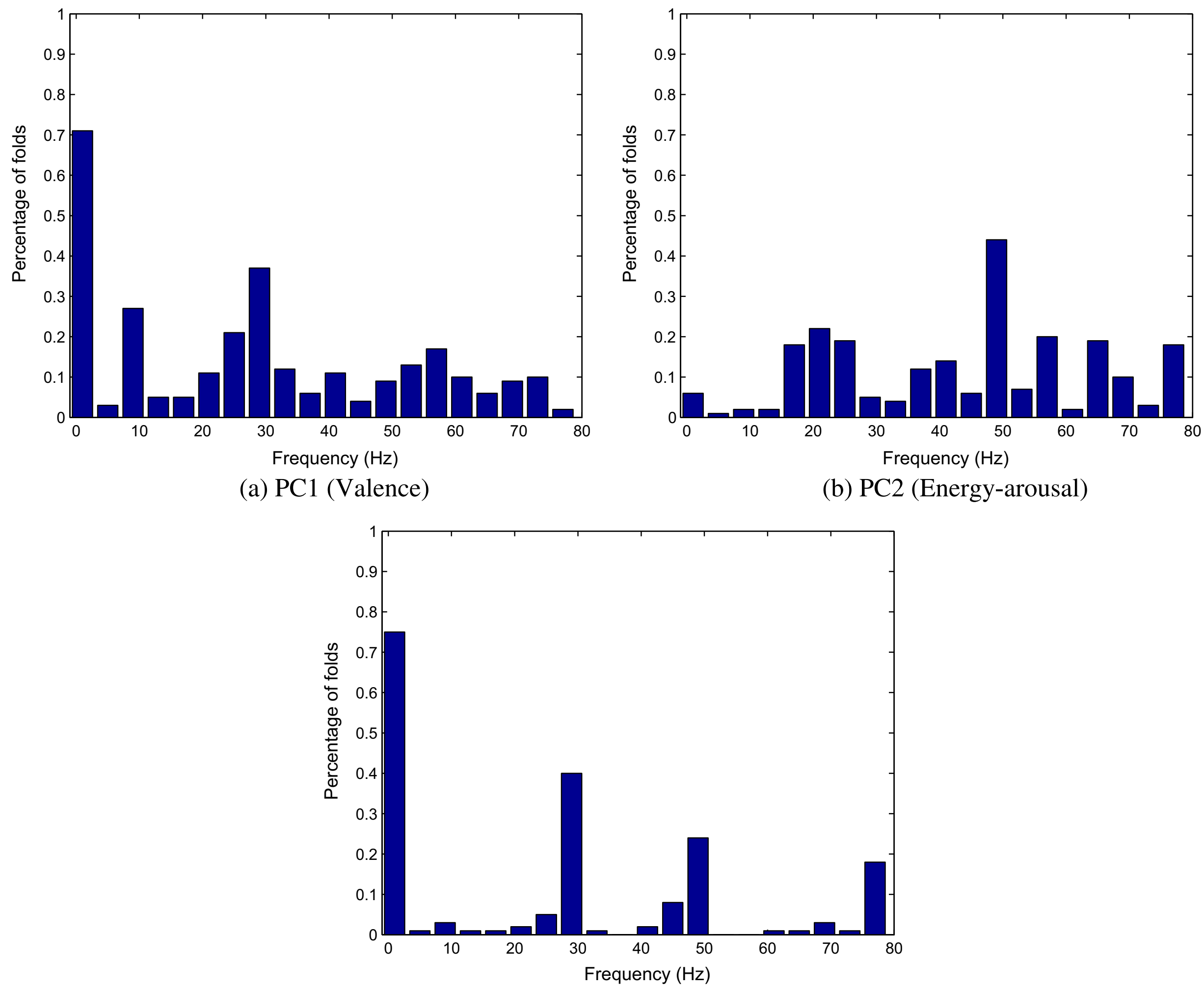
Fig. 2. The sub-set of features selected for prediction of music-induced emotions and the percentage of folds in which they were selected. Features are illustrated if they are selected in more than 5% of the folds of the cross-validation procedure. The blue bars indicate features derived from the EEG, while the red bars indicate music-derived features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The observation of the involvement of the left motor cortex in the prediction of energy-arousal is interesting. Previous work has shown that changes in the tempo of a piece of music entrains activity in the left motor cortex (Daly et al., 2014; Daly et al., 2014). It has also been reported that faster music tempo can increase arousal without affecting mood (Husain, Thompson, & Schellenberg, 2002). It could be the case that changes in the tempo of the music produce changes in activity over the left motor cortex, which correlates with changes in arousal and, therefore, can be used as a feature to predict music-induced energy-arousal. Further to this, it has been reported that music tempo preference is correlated with the beta rhythm in the EEG (Bauer et al., 2015), which is partly supported by our finding that the beta rhythm is involved in valence, a measure of how pleasant or unpleasant a listener finds the music.

Tension-arousal (response PC3) is observed to be predicted by band-powers in the delta, beta, and gamma frequency bands. These band-powers are observed to lie over the right motor cortex and the parietal cortex. Additionally, alpha asymmetry is observed to be a very good predictor of tension-arousal.

These results correspond well to our previous work. For example, in Daly et al. (2014) we report a pattern of connectivity between regions of the right motor cortex that correlates with changes in music-induced tension.

These results also correspond well to previously reported work in the literature. Specifically, the different emotional responses to music each produce distinctly different effects within the EEG over different spatial locations. This further reinforces the view that there is no single pathway for all musical emotions, but rather that there are distinct networks of brain regions involved in different emotions (Peretz, 2009, chap. 5). These networks are also likely to involve sub-cortical regions, however, from EEG analysis alone this cannot be verified.



(a) PC1 (Valence)

(b) PC2 (Energy-arousal)

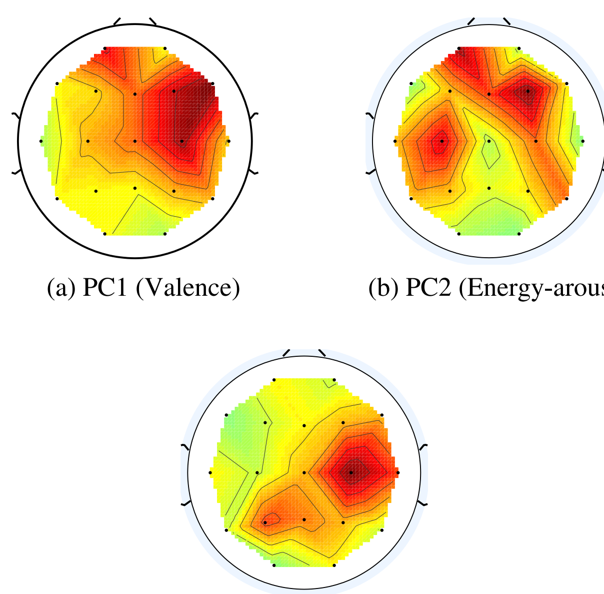
(c) PC3 (Tension-arousal)

Fig. 3. Histograms of the frequency bands from which band-power features are selected for use in the regression models used to predict emotional responses to music.

Thus, the regression models we identify in this study are able to accurately predict the affective responses of listeners to previously unheard pieces of music. The correlation between the predicted affective responses and the actual responses is highest for valence. This may be due to the selection of the candidate EEG features and acoustic features available in this study more strongly relating to valence than arousal or the linear regression models better modelling relationships between our candidate features and valence. It may also be the case that valence is a more stable, less variable, and hence more easily predicted affective response to music. However, this would need further verification.

The relationships observed in this study are correlational in nature and indicate which acoustic properties of a variety of pieces of music relate to perceived emotions in a population of 31 listeners. However, the complex nature of these interactions means that it is difficult to determine the specific relationships between these features.

Thus, if one wishes to modulate the emotions induced by listening to a piece of music, it is not necessarily simply a matter of manipulating one acoustic feature of the music independently of the others. Rather, both the relationships between the music features and the listener's current neurophysiological state needs to be understood before the emotional response of a listener to a particular piece of music can be predicted.

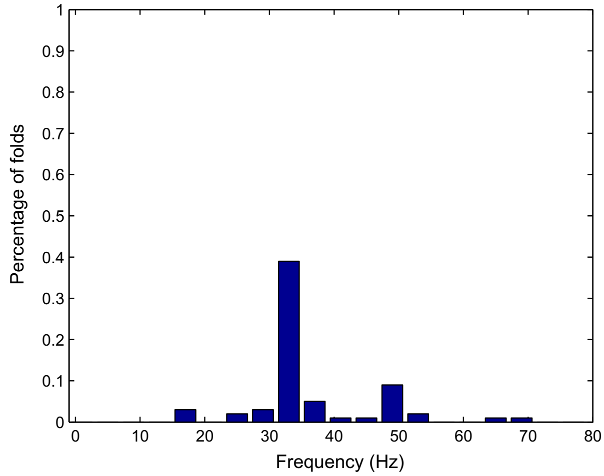


(a) PC1 (Valence)

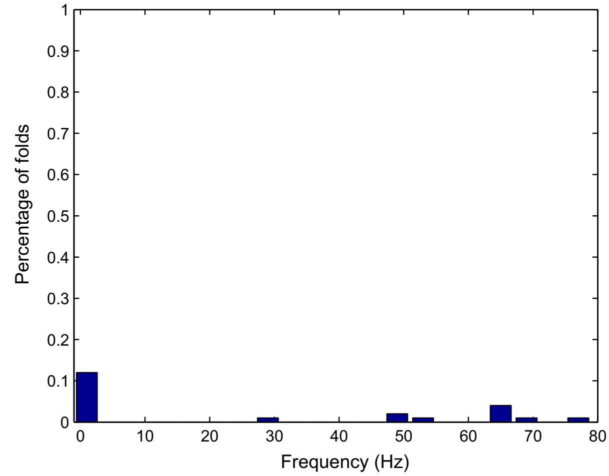
(b) PC2 (Energy-arousal)

(c) PC3 (Tension-arousal)

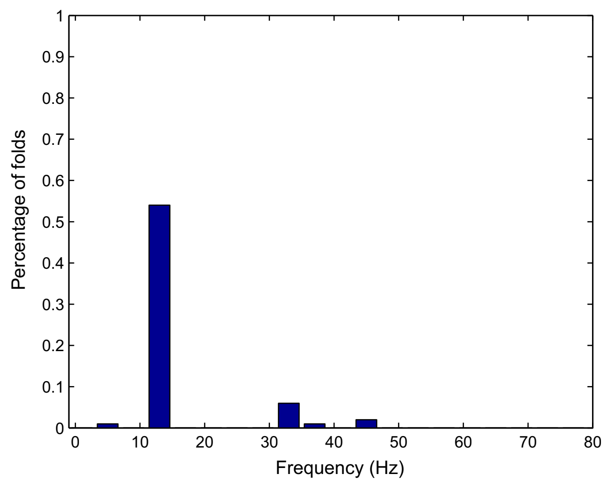
Fig. 4. Scalp maps of the spatial locations from which band-power features are selected for use in the regression models used to predict emotional responses to music.



(a) PC1 (Valence)



(b) PC2 (Energy-arousal)



(c) PC3 (Tension-arousal)

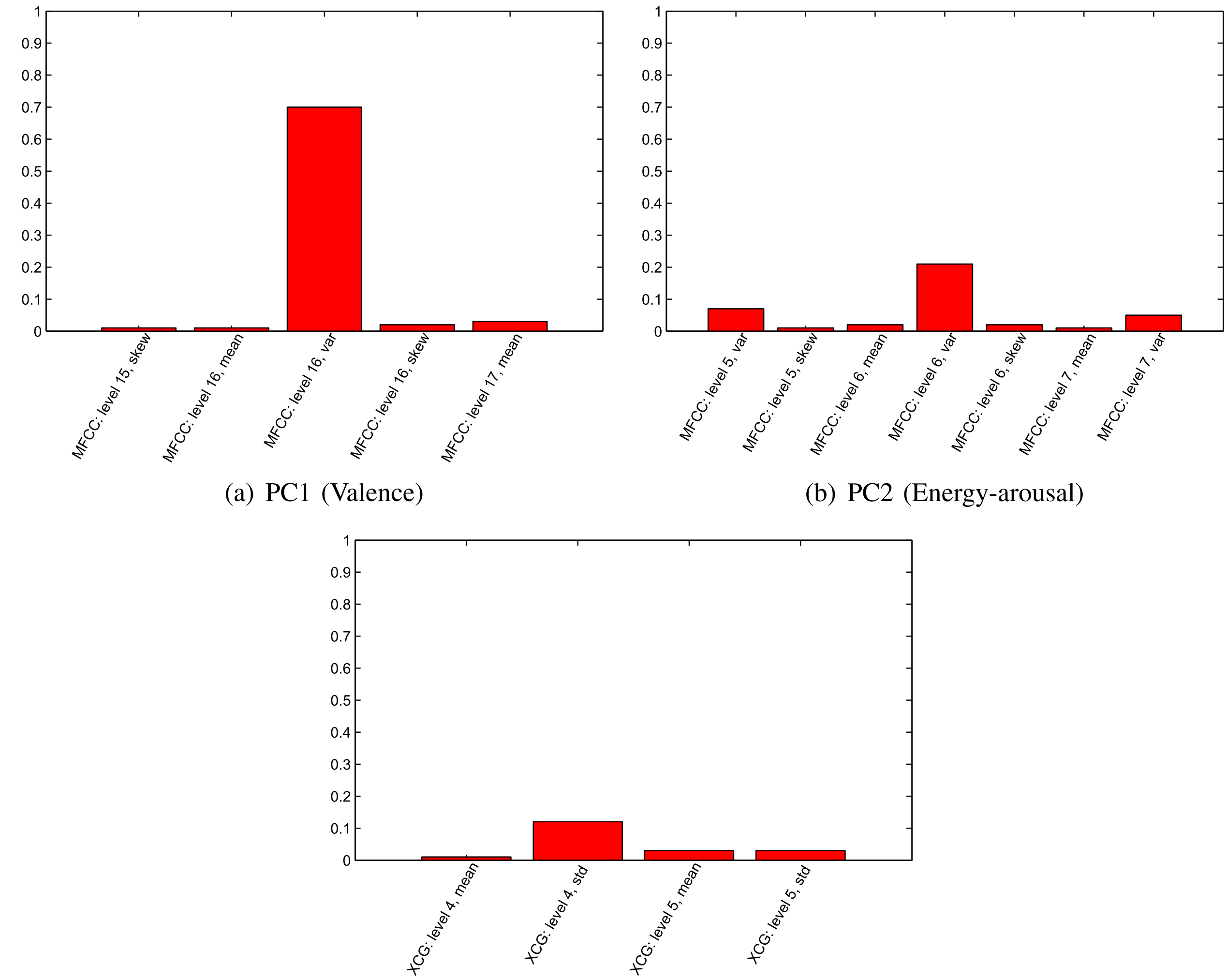
Fig. 5. Histograms of the frequency bands from which asymmetry features are selected for use in the regression models used to predict emotional responses to music. Note, bands have a width of 4 Hz.

The acoustic features we use are estimated directly from the music signals via the use of signal processing methods. Therefore, they provide a way of examining the music from a signal processing perspective that is independent of any specific music theory. However, consequently the translation between this perspective and that provided by specific music theories composed of notes, keys, and scores is not immediately apparent for the majority of acoustic features. These musical features tend to be structural or higher-level combinations of various acoustic features. Relating the acoustic features we identify as suitable for use in predicting music-induced emotion to their corresponding musical features is a significant area for further work, which is outside the scope of this study.

The results reported in this study provide some reinforcement for findings reported elsewhere that some music properties, such as tempo, relate to affective states based upon energy and arousal (Husain et al., 2002). They also support findings that valence perceived by listening to music relates to EEG measures including prefrontal asymmetry.

Finally, our findings may also be compared to results reported in Livingstone and Thompson (2006) in which relationships between musical descriptors and perceived emotions are described. In Livingstone and Thompson (2006) pitch is reported to weakly correlate with valence, a finding also supported by our work.

Future work will seek to build upon these findings to construct a generative music system to create novel music containing particular combinations of acoustic properties known to induce particular emotional states. This has applications in, amongst other areas, music therapy, and the emerging field of brain-computer music interfacing (Daly et al., 2014; Miranda, Magee, Wilson, Eaton, & Palaniappan, 2011).



(a) PC1 (Valence)

(b) PC2 (Energy-arousal)

(c) PC3 (Tension-arousal)

Fig. 6. Acoustic features selected for prediction of the response PCs.

Acknowledgment

This work was supported by the EPSRC grants EP/J003077/1 and EP/J002135/1.

References

Bauer, A.-K. R., Kreutz, G., & Herrmann, C. S. (2015). Individual musical tempo preference correlates with EEG beta rhythm. *Psychophysiology, 52*(4), 600–604. <http://www.ncbi.nlm.nih.gov/pubmed/25353087>.

Bradt, J., Magee, W. L., Dileo, C., Wheeler, B. L., & McGilloway, E. (2010). Music therapy for acquired brain injury. *The Cochrane Database of Systematic Reviews* (7), CD006787. <http://www.ncbi.nlm.nih.gov/pubmed/20614449>.

Chen, X. (1992). Active music therapy for senile depression. *Zhonghua shen jing jing shen ke za zhi = Chinese Journal of Neurology and Psychiatry, 25*(4), 208–210. 252–3, <http://www.ncbi.nlm.nih.gov/pubmed/1478135>.

Craig, D. G. (2005). An exploratory study of physiological changes during chills induced by music. *Musicae Scientiae, 9*(2), 273–287. <http://msx.sagepub.com/content/9/2/273.abstract>.

Daly, I., Pichiorri, F., Faller, J., Kaiser, V., Kreilinger, A., Scherer, R., & Mueller-Putz, G. (2012). What does clean EEG look like? In *Conf proc IEEE eng med biol soc*.

Daly, I., Hallowell, J., Hwang, F., Kirke, A., Malik, A., Roesch, E., ... Nasuto, J. (2014). Changes in music tempo entrain movement related brain activity. In *Proceedings of the EMBC*.

Daly, I., Hwang, F., Kirke, A., Malik, A., Weaver, J., Williams, D., ... Nasuto, S. J. (2014). Automated identification of neural correlates of continuous variables. *Journal of Neuroscience Methods*. <http://www.sciencedirect.com/science/article/pii/S0165027014004336>.

Daly, I., Malik, A., Hwang, F., Roesch, E., Weaver, J., Kirke, A., ... Nasuto, S. J. (2014). Neural correlates of emotional responses to music: An EEG study. *Neuroscience Letters, 573*, 52–57.

Daly, I., Williams, D., Hwang, F., Kirke, A., Malik, A., Roesch, E., ... Nasuto, S. J. (2014). Investigating music tempo as a feedback mechanism for closed-loop BCI control. *Brain-Computer Interfaces*, 1–12. <http://www.tandfonline.com/doi/abs/10.1080/2326263X.2014.979728#.VJGE-DGsWQA>.

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 28*(4), 357–366.

Dubnov, S. (2006). CATbox: Computer audition toolbox in matlab. (p. 1). <http://musicweb.ucsd.edu/sdubnov/CATbox/CATbox.htm>.

Eerola, T., & Vuoskoski, J. K. (2010). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*(1), 18–49. <http://pom.sagepub.com/content/39/1/18.abstrac>.

Ellis, D. P. W. (2007). Beat tracking by dynamic programming. *Journal of New Music Research, 36*(1), 51–60. <http://dx.doi.org/10.1080/09298210701653344>.

Erkkilä, J., Punkanen, M., Fachner, J., Ala-Ruona, E., Pöntiö, I., Tervaniemi, M., ... Gold, C. (2011). Individual music therapy for depression: Randomised controlled trial. *The British Journal of Psychiatry: The Journal of Mental Science, 199*(2), 132–139. <http://bjp.rcpsych.org/content/199/2/132.short>.

Etzel, J. A., Johnsen, E. L., Dickerson, J., Tranel, D., & Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 61*(1), 57–69. <http://www.ncbi.nlm.nih.gov/pubmed/16460823>.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and models*. Berlin: Springer.

Flores-Gutiérrez, E. O., Díaz, J.-L., Barrios, F. A., Favila-Humara, R., Guevara, M. A., del Río-Portilla, Y., et al. (2007). Metabolic and electric brain patterns during pleasant and unpleasant emotions induced by music masterpieces. *International Journal of Psychophysiology: Official Journal of the International Organization of*

Psychophysiology, 65(1), 69–84. <http://www.ncbi.nlm.nih.gov/pubmed/17466401>.

Gabrielsson, A. & Juslin, P. N. (2003). Emotional expression in music. In *Handbook of affective sciences* (pp. 503–534).

Gabrielsson, A. & Lindström, E. (2001). The influence of musical structure on emotional expression. In *Music and emotion: Theory and research* (pp. 223–248).

Gosselin, N. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain, 128*(3), 628–640. <http://www.ncbi.nlm.nih.gov/pubmed/15699060>.

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia, 45*(2), 236–244. <http://www.ncbi.nlm.nih.gov/pubmed/16970965>.

Goto, M. (2006). A chorus section detection method for musical audio signals and its application to a music listening station. *IEEE Transactions on Audio, Speech and Language Processing, 14*(5), 1783–1794. <http://dl.acm.org/citation.cfm?id=2209815.2210582>.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, 4*(1), 47–56.

Husain, G., Thompson, W., & Schellenberg, E. (2002). Effects of musical empo and mode on arousal, mood, and spatial abilities. *Music Perception, 20*, 151–171.

Jausovec, N., & Jausovec, K. (2005). Differences in induced gamma and upper alpha oscillations in the human brain related to verbal/performance and emotional intelligence. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 56*(3), 223–235. <http://www.sciencedirect.com/science/article/pii/S0167876004002284>.

Kedem, B. (1986). Spectral analysis and discrimination by zero-crossings. *Proceedings of the IEEE, 74*(11), 1477–1493.

Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(12), 2067–2083. <http://www.ncbi.nlm.nih.gov/pubmed/18988943>.

Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R., et al. (2010). EEG-based emotion recognition in music listening. *IEEE Transactions on Bio-Medical Engineering, 57*(7), 1798–1806. <http://www.ncbi.nlm.nih.gov/pubmed/20442037>.

Li, T., & Ogihara, M. (2004). Music artist style identification by semi-supervised learning from both lyrics and content. In *Proceedings of the 12th annual ACM international conference on Multimedia – MULTIMEDIA '04* (pp. 364). New York, New York, USA: ACM Press.

Li, T., Ogihara, M., & Li, Q. (2003). A comparative study on content-based music genre classification. In *Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval – SIGIR '03* (pp. 282). New York, New York, USA: ACM Press.

Livingstone, S. R., & Thompson, W. F. (2006). Multimodal affective interaction. *Music Perception, 24*(1), 89–94. <http://espace.library.uq.edu.au/view/UQ:8039>.

Livingstone, R. S., & Thompson, W. F. (2009). The emergence of music from the theory of mind. *Musicae Scientiae, 13*(2 Suppl), 83–115. <http://msx.sagepub.com/content/13/2_suppl/83.abstract>.

Lundqvist, L.-O., Carlsson, F., Hilmersson, P., & Juslin, P. N. (2008). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music, 37*(1), 61–90. <http://pom.sagepub.com/content/early/2008/10/15/0305735607086048.short>.

Maratos, A. S., Gold, C., Wang, X., & Crawford, M. J. (2008). Music therapy for depression. *The Cochrane Database of Systematic Reviews* (1), CD004517. <http://www.ncbi.nlm.nih.gov/pubmed/18254052>.

Martin Mckinney, J. B. (2003). Features for audio and music classification. In *Proceedings of the international symposium on music information retrieval*.

McDermott, O., Crellin, N., Ridder, H. M., & Orrell, M. (2013). Music therapy in dementia: A narrative synthesis systematic review. *International Journal of Geriatric Psychiatry, 28*(8), 781–794. <http://www.ncbi.nlm.nih.gov/pubmed/23080214>.

Miranda, E. R., Magee, W. L., Wilson, J. J., Eaton, J., & Palaniappan, R. (2011). Brain-Computer Music Interfacing (BCMI): From basic research to the real world of special needs. *Music and Medicine, 3*(3), 134–140.

Mitrovic, D., Zeppelzauer, M., & Breiteneder, C. (2006). Discrimination and retrieval of animal sounds. In *2006 12th International multi-media modelling conference. IEEE* (pp. 339–343).

Mitrović, D., Zeppelzauer, M., & Breiteneder, C. (2010). Features for content-based audio retrieval. In *Advances in computers* (Vol. 78, pp. 71–150).

Morchen, F., Ultsch, A., Thies, M., & Lohken, I. (2006). Modeling timbre distance with temporal statistics from polyphonic music. *IEEE Transactions on Audio, Speech and Language Processing, 14*(1), 81–90.

Pampalk, E., Rauber, A., & Merkl, D. (2002). Content-based organization and visualization of music archives. In *Proceedings of the tenth ACM international conference on Multimedia – MULTIMEDIA '02* (pp. 570). New York, New York, USA: ACM Press.

Peretz, I. (2009). Towards a neurobiology of musical emotions – Oxford scholarship. *Handbook of Music and Emotion: Theory, Research, Applications*, 99–126. <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199230143.001.0001/acprof-9780199230143-chapter-5>.

Peretz, I., Aube, W., & Armony, J. L. (2013). Toward a neurobiology of musical emotions. In *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man* (pp. 277–299). Oxford University Press. <http://books.google.com/books?hl=en&lr=&id=EQvmZn1Ino4C&pgis=1>.

Rabiner (1979). *Digital processing of speech signals*. Pearson Education. <http://books.google.com/books?id=JAIj5fucWiUC&pgis=1>.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 1161–1178.

Scheirer, E., & Slaney, M. (1997). Construction and evaluation of a robust multifeature speech/music discriminator. *1997 IEEE international conference on acoustics, speech, and signal processing* (Vol. 2, pp. 1331–1334). IEEE Comput. Soc. Press.

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research, 33*(3), 239–251.

Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality, 14*(4), 21.

Schmidt, L. A., & Trainor, L. J. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion, 15*(4), 487–500. <http://dx.doi.org/10.1080/02699930126048>.

Schomer, D. L., Blum, A. S., & Rutkove, S. B. (2007). The clinical neurophysiology primer. In A. S. Blum & S. B. Rutkove (Eds.). Totowa, NJ: Humana Press. <http://www.springerlink.com/content/m6v44115r6803181/>.

Schubert, E. (1999). *Measurement and time series analysis of emotion in music*. <http://philpapers.org/rec/SCHMAT-13>.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America, 36*(12), 2346.

Silberman, E. K., & Weingartner, H. (1986). Hemispheric lateralization of functions related to emotion. *Brain and Cognition, 5*(3), 322–353. <http://www.sciencedirect.com/science/article/pii/0278262686900357>.

Stevens, S. S. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America, 8*(3), 185.

Tamplin, J., & Baker, F. (2006). *Music therapy methods in neurorehabilitation: A clinician's Manual*. Jessica Kingsley Publishers. <http://books.google.com/books?hl=en&lr=&id=iIu3rAsairEC&pgis=1>.

Thompson, W., & Robitaille, B. (1992). Can composers express emotions through music? *Empirical Studies of the Arts, 10*, 79–89.

van Tricht, M. J., Smeding, H. M. M., Speelman, J. D., & Schmand, B. A. (2010). Impaired emotion recognition in music in Parkinson's disease. *Brain and Cognition, 74*(1), 58–65. <http://www.sciencedirect.com/science/article/pii/S027826261000076X>.

Walpulski, M. (2008). EEG representation of emotion evoking pictures. Tech. Rep., <http://essay.utwente.nl/58961/1/scriptie_M_Walpsuki.pdf>.

Wilkins, R. W., Hodges, D. A., Laurienti, P. J., Steen, M., & Burdette, J. H. (2014). Network science and the effects of music preference on functional brain connectivity: From Beethoven to Eminem. *Scientific Reports, 4*, 6130. <http://www.nature.com/srep/2014/140828/srep06130/full/srep06130.html>.

Zhang, T., & Kuo, C. (2001). *Content-based audio classification and retrieval for audiovisual data parsing*. Springer.