

Using a Combination of Electroencephalographic and Acoustic Features to Accurately Predict Emotional Responses to Music

Devon Krish

Los Gatos High School, Los Gatos, CA, USA

ABSTRACT

Music has the ability to evoke a wide variety of emotions in human listeners. Research has shown that treatment for depression and mental health disorders is significantly more effective when it is complemented by music therapy. However, because each human experiences music-induced emotions differently, there is no systematic way to accurately predict how people will respond to different types of music at an individual level.

In this experiment, a model is created to predict humans' emotional responses to music from both their electroencephalographic data (EEG) and the acoustic features of the music. By using recursive feature elimination (RFE) to extract the most relevant and performing features from the EEG and music, a regression model is fit and accurately correlates the patient's actual music-induced emotional responses and model's predicted responses. By reaching a mean correlation of $r = 0.788$, this model is significantly more accurate than previous works attempting to predict music-induced emotions (e.g. a 370% increase in accuracy as compared to [Daly et al. \(2015\)](#)).

The results of this regression fit suggest that accurately predicting how people respond to music from brain activity is possible. Furthermore, by testing this model on specific features extracted from any musical clip, music that is most likely to evoke a happier and pleasant emotional state in an individual can be determined. This may allow music therapy practitioners, as well as music-listeners more broadly, to select music that will improve mood and mental health.

Keywords: EEG, music therapy, acoustic features, machine learning, emotional-response predictions

I. INTRODUCTION

Music is known to be an extremely powerful tool that can make listeners feel pleasure, happiness, sadness, and even fear ([Fritz et al. 2009](#)). Music therapy is also a health-intervention that has subsequently proven to be an effective treatment for poor mental health, mood disorders, and depression ([Maratos et al. 2008](#)).

For example, music therapy was determined to significantly improve mood when compared to treatment as usual ([Maratos et al. 2008](#); [Ramirez et al. 2018](#)). Furthermore, [Chen \(1992\)](#) determined antidepressant drugs coupled with music therapy were more effective than antidepressant drugs alone in

improving mental health. Also, [Koelsch & Jäncke \(2015\)](#) concluded that music can reduce pain and anxiety in patients with heart disease by lowering heart rate and blood pressure.

In music therapy, a therapist prescribes music selections to a patient to listen to based solely on the therapist's expertise, experiences, and evaluation of the patient ([Tamplin & Baker 2006](#); [Maratos et al. 2008](#)). However, to determine the optimal music to prescribe to the patient, the individual's emotional reaction to the signal must be predicted. Predicting a human's emotional response to a musical selection they have never heard is an extreme challenge because all humans experience music-induced emotions differently depending on their life experiences, moods, influences,

gender (Hunter, Schellenberg & Schimmack 2010; McRae et al. 2008), and a wide variety of other factors. Despite music therapy's promising results as a healing agent for depression and mental health disorders, music therapy does not currently rely on a systematic method to predict emotional responses to music on an individual level.

Many models for the relationship between induced emotions in humans and attributes of music have been created. For example, Schubert (2004) studied the effects of musical features and music theory attributes such as dynamics and melodic contour on emotions. Gabrielsson & Lindström (2001) subsequently studied the effects of variations of tempo, articulation, dynamics, and intonation on perceived emotion by using a model in which emotional responses ranging from strong to weak were plotted across two axes: one for pleasantness and one for excitement. This model also featured plots of tempo and other musical descriptors. This model may provide information about how individual music descriptors affect emotional responses. However, because all people perceive emotions differently and therefore individual-level emotional predictions are necessary, using physiological data may be of significant value.

Other researchers have used physiological measurements as correlates for induced emotional responses. For example, Etzel et al. (2006) tested the effect of music meant to evoke different moods including happiness, sadness, and fear on cardiovascular activity. Similar research has found that certain selections of pleasant and happiness-inducing music may increase heart rate (Brouwer et al. 2013), as revealed through the analysis of electrocardiogram (ECG) signals (Kim & André 2008). More physiological measurement techniques used to correlate music-induced emotions include skin conductance response (SCR) (Khalfa, Isabelle, Jean-Pierre, & Mannon 2002) electromyography (EMG) for facial expression (Lundqvist et al. 2008), electrodermal activity (EDA) (Craig 2005), respiration rate (Merrill et al. 2020), and pupillometry (Nakakoga et al. 2020).

On the other hand, researchers have also turned to studying brain activity to measure music-induced

emotions. Such experiments have been performed using functional magnetic resonance imaging (fMRI) for example (Koelsch et al. 2006; Brattico et al. 2011). Alternatively, researchers have also used electroencephalographic (EEG) data (Daly et al. 2015; Schmidt & Trainor 2001).

Combining EEG with music, video, and other media features has proven to be effective at predicting emotional responses (Kortelainen & Seppänen 2013; Li et al. 2018) because music and their induced emotions are thoroughly believed to engage large networks of neurons and neuronal structures (Daly et al. 2019). For instance, a study by van Tricht et al. (2010) found that Parkinson's disease impaired the emotional recognition of fear and anger in music. Similarly, the deracination of the anteromedial temporal lobe impaired the music-induced recognition of terror and scariness. (Gosselin et al. 2011). Thus, music-induced emotions may relate to a variety of changes in EEG.

That being said, using EEG to predict music-induced emotions has shown to be a difficult feat because EEG is non-stationary and very noisy (Daly et al. 2015; Koelstra et al 2012; Jiang et al. 2019). This problem is further challenged by combining EEG with variations in how humans perceive emotions (Barrett et al. 2007), differences in musical preferences (Bauer, Kreutz, & Hermann 2015), and disparities in age and gender which may influence these preferences (Vieillard & Gilet 2013; Le Blanc et al. 1999).

Therefore, by selecting only the most important and relevant descriptors, a combination of both acoustic and EEG features may be used to train a model to predict music-induced emotions at a high accuracy on the individual level.

Since classical-style music has been shown to induce emotions at a strong level (Schaefer 2017; Kreutz et al. 2007), this study utilizes the dataset created by Daly et al. (2015), in which classical music clips are played to participants while their EEG is recorded. Descriptive features from the musical clips (from the set of stimuli from Eerola & Vuoskoski 2010) are then extracted and brain activity to train a Lasso regression model to predict each participant's emotional responses to music.

II. MATERIALS AND METHODS

2.1 Methods and Experimental Data

This analysis utilized the EEG and emotional response data gathered from [Daly et. al \(2015\)](#). The data consisted of thirty-one individuals, thirteen males and eighteen females, whose ages ranged from 18 to 66. Each participant's electroencephalogram (EEG) was recorded from 19 electrode channels positioned from the nasion to inion as according to the 10-20 EEG Placement System ([Milnik 2006](#)).

This study also featured classical music stimuli drawn from a dataset of 360 excerpts from film scores including titles such as *Psycho (1960)*, *Gladiator (2000)*, and *Big Fish (2003)*. The musical stimuli in this dataset were specifically chosen to induce specific emotions in human listeners ([Eerola & Vuoskoski 2010](#)).

The participants were told to stay still and each listened to 40 randomly drawn musical selections from [Eerola & Vuoskoski \(2010\)](#) for 15 seconds. Immediately after, they responded to a series of eight Likert-scale questions on a scale of strongly disagree to strongly agree to identify their emotional response across eight axes: happiness, sadness, pleasantness, fear, anger, tenderness, tension, and energy-level ([Daly et al. 2015](#)).

Because some of these emotion categories are likely to be highly correlated ([Larsen & McGraw 2001](#)), a Principal-Component Analysis (PCA) was used to reduce the eight axes into three principal components (PCs), which explains a large amount of the variance (75%) of the participant's music-induced emotional responses ([Daly et al. 2015](#)). These three PCs represent valence-arousal, a measurement of pleasure and happiness; energy-arousal, a quantification of liveliness; and tension-arousal, a measurement of tenseness ([Ilie & Thompson 2006](#); [Ilie & Thompson 2011](#)), as according to the Schimmack and Grob model ([Schimmack & Grob 2000](#)). It is these three PCs that are ultimately subjected to further analysis.

Acoustic features from the music played to each participant from the musical dataset and various

anatomical features from the participants' EEG data are then extracted. Through recursive feature elimination (RFE), the most performing features are selected to accurately predict a participant's self-reported emotional response to music along each of the three axes of the PCs.

2.2 EEG Features

Because EEG signals are so noisy ([Daly et al. 2012](#); [Jiang et al. 2019](#)), the EEG signals were first pre-processed to remove artefacts and noise. Discrete Wavelet Transform (DWT) was then used to decompose the EEG signals into alpha (α) [12-24Hz), beta (β) [24-48Hz), gamma (γ) [48-80Hz), delta (δ) (0-6Hz), and theta (θ) [6-12Hz) wavelet bands. DWT provides high-frequency resolution at the high and low EEG frequencies. DWT is an extremely convenient tool for processing non-stationary EEG and rendering these signals more suitable for feature extraction than just standard EEG alone ([Qazi et al. 2016](#)). From the α , β , γ , δ , and θ wavelet bands, a total of 285 features were extracted from the EEG signals using Minimum Norm Estimates (MNE). RFE was further utilized to determine that the features below were beneficial to training the model.

The energy of each wavelet band, or strength of the signal at any time interval as it represents the area under the curve (AUC), was computed.

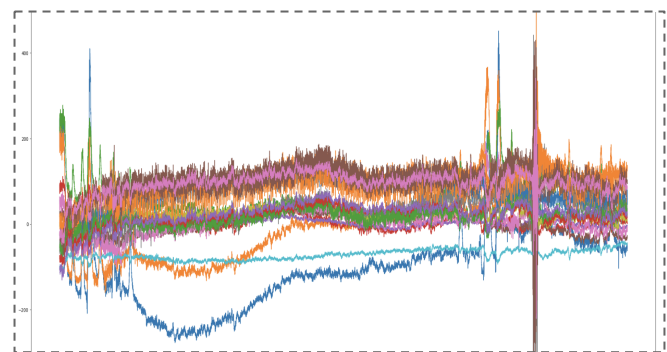


Fig. 1: A visual depiction of the energy of the EEG signals for each of the 19 electrodes for one participant, where the strength is represented by the y-axis and time is represented by the x-axis (15 seconds, or one song)

The mean and standard deviation of each wavelet band was also extracted and used to train the model. The standard deviation supplies information about how close the features are to the signal's mean and is mathematically represented as shown below:

$$STD = \sqrt{\sum_{i=1}^N \frac{(x_i - \mu)^2}{N-1}} \quad (1)$$

Where x_i represents the random variable with a mean signal of μ

The entropy of an EEG signal represents the outcome uncertainty measurement and is calculated through the following formula:

$$H(x) = - \sum_{i=1}^n P(x_i) \log[P(x_i)] \quad (2)$$

Where x represents the random variable with possible outcomes $x(i=1), \dots, x_n$, which occur with a probability $P(x_i)$

The Hjorth parameter of mobility is a normalized slope descriptor that represents the square root of the variance of the first derivative of the signal divided by the variance of the signal (Nascimben, Ramsøy, Bruni 2019). It is otherwise stated as the proportion of standard deviation to a signal's power spectrum and has been noted to be advantageous to emotion recognition (Li et al. 2018). Results from the addition of the Hjorth parameter to the model significantly improved the model's accuracy, as hypothesized. This mobility parameter is mathematically represented as shown below:

$$H_m(t) = \sqrt{\frac{\text{var}\left(\frac{dy(t)}{dt}\right)}{\text{var}(y(t))}} \quad (3)$$

$H_m(t)$ equates to the square root of the variance of the first derivative of signal $y(t)$ divided by the total variance of signal $y(t)$

Power frequency bands extract features by using frequency bands to compute the power spectrum of an EEG signal (Al-Fahoum & Al-Fraihat 2013). The five frequency wavelet bands extracted the power of

each of the 19 electrodes' signals, resulting in 95 features.

2.3 Acoustic Features

From each of the 360 musical clips used as stimuli, a range of specific acoustic features was extracted. These extracted features included both spectral and temporal types, ranging from describing each piece's key and tonal qualities to the speed. A total of 352 musical features were selected. RFE was also used to make sure the acoustic features were beneficial to the model. Furthermore, for each of the eleven acoustic feature types (except key and bpm), the kurtosis, maximum, mean, median, minimum, skewness, and standard deviation were also extracted. The acoustic feature types are described below.

Chroma and chromagrams can be described as the transformation of a musical signal's pitches (twelve possible pitches in total) through time. From chroma, both the intensity and certain pitch are extracted - a total of 84 features.

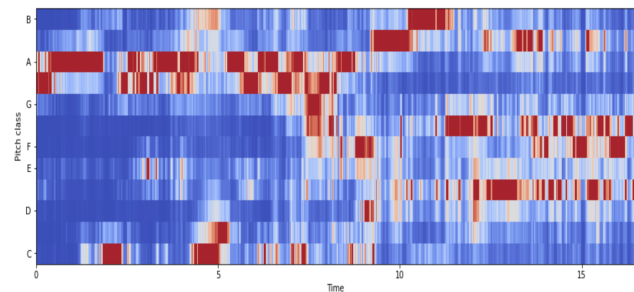


Fig. 2: A chromagram in which the x-axis is represented by time in seconds of a signal while the y-axis represents the pitches. The more red a pitch is at a certain time, the more intense that pitch is.

BPM or beats per minute provides the number of beats in the signal per minute. A low tempo of 40 BPM, for example, is very likely to indicate a slower, solemn musical stimulus while a tempo of 160 BPM is likely to indicate an upbeat, faster-paced musical piece. BPM proved to be an extremely important feature in determining the emotional responses to a musical selection.

Zero-crossing rate (ZCR) can be described as the rate at which the signal of a music stimulus may change from positive to negative or negative to positive,

thus crossing zero (zero-crossings) in a fixed amount of time; ZCR amounted to a total of seven features.

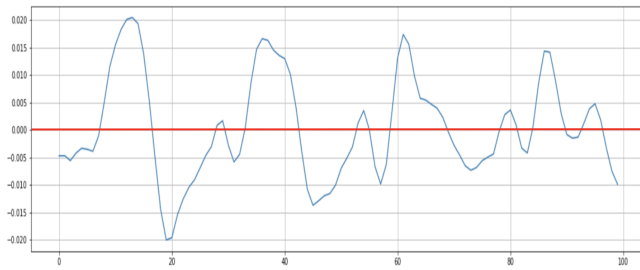


Fig. 3: A plotted depiction of ZCR in which every time the signal crosses $y = 0$ (illustrated by the red line), the number of zero-crossings (n) increases by one.

Major and minor keys turned out to be an extremely significant feature in determining the emotional responses to an audio signal. A major key versus a minor key is determined by analyzing the notes or pitches present in a signal. It is also widely acknowledged that songs in major keys tend to sound bright and cheerful while songs in minor keys are more melancholy.

MFCC or Mel-frequency cepstral coefficients are computed by taking the logarithm of the Fourier coefficients of an audio signal that has been converted to the Mel-scale. MFCCs ultimately represent the timbre of the audio signal. (Stevens 1937; Garima & Barkha 2013). These 140 features are also commonly used in speech recognition systems as detailed in (Anggraeni et al. 2018).

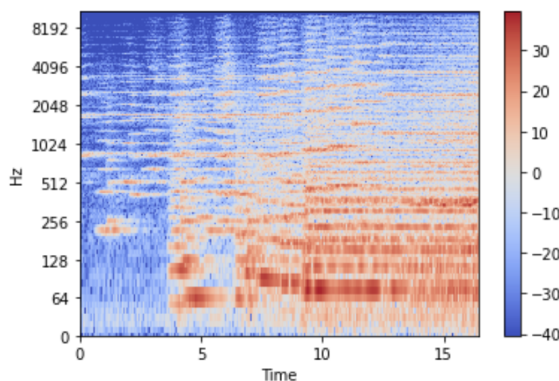


Fig. 4: An MFCC spectrum plot in which the x-axis represents the time in seconds of a signal and the y-axis represents the increasing MFCC coefficients of a signal. Similar to the chromagram in Fig. 1, the redder the pitch is at a certain time, the more intense that coefficient is.

Tonnetz is a feature that determines the tonal centroids, or harmonic components of the signal when extracted.

The spectral centroid of a musical signal defines which frequency the energy of a spectrum is centered upon or where the center of mass of the spectrum is located. The spectral roll-off is the frequency below which a percentage (normally 0.85) of the spectral energy of the signal lies (Jang et al. 2008). The spectral contrast is the difference in level between the crest and troughs of the spectrum. The spectral flatness can determine how noisy a signal is in decibels (with 1.0 indicating the spectrum is white noise). Finally, the spectral bandwidth, or the extent of the power transfer around the center frequency of the audio signal, is extracted (Theimer et al. 2008). Altogether, the spectral features contribute 77 features to the model.

2.5 Training the Model

A total of 637 extracted features (285 EEG features and 352 acoustic features) were used to train a Lasso-regularized linear regression model (with an alpha parameter set to 10) to predict the participants' emotional responses to music in terms of the PCs. To best fit the data and generate accurate predictions, a multi-task, cross-validated, Lasso regression model with five folds was used. In each of the five folds, the features were split into a training and testing set. Combinations of the training set features were then related to the PCs to create a linear regression model that would fit the emotional responses of the participants as noted by their recorded PCs.

After being trained on the said training set, the model attempted to predict the participant response PCs from the previously unseen features in the testing set. By identifying how close the model's predicted response PCs are to the participants' actual, recorded PCs for each data point in the testing set of each cross-fold, the model's performance and statistical can be determined.

2.4 Recursive Feature Elimination (RFE)

As first utilized for gene selection, RFE is an effective method that determines the importance of features. The RFE algorithm was trained on the whole set of features initially (637 features: 285 from the EEG and 352 from the music). The algorithm could then determine the importance of each feature to the correlation model by assigning weights and eliminating the lowest-ranking features (Li et al. 2018). This process occurred recursively for several rounds until all the features had been selected. Using an RFECV (cross-validation) selector, the 17 feature types (comprising the 637 features) were ranked to determine which features were most beneficial to obtain the regression model's correlation. RFE was also used to determine that the features would only improve the model's prediction accuracy.

III. RESULTS

3.1 Correlation Analysis

The model's prediction performance is first evaluated on the EEG and acoustic features separately and then combined. These feature subsets are then used to evaluate the response PCs individually. For each PC, the mean correlation (r), between the actual recorded PC and the predicted PC is calculated. The results of these correlations are displayed in Table 1. Given the noise and non-stationarity of the EEG data (Hassani & Karami 2015) and the 75% variance of the participant's emotional responses (Daly et al. 2015), training the model on each feature subset resulted in a predicted response PC of high correlation ($p < 0.001$). As displayed in Table 1, the model's correlation was highest when both the EEG and acoustic features are combined.

The results of this correlation are unparalleled to previous works dealing with predicting music-induced emotions and EEG. For example, our r -values are almost 370% better than a comparable study by Daly et al. (2015).

Response PCs	Mean correlation (r)		
	Both	Acoustic	EEG
PC1 (Valence-arousal)	0.774	0.474	0.355
PC2 (Tension-arousal)	0.791	0.362	0.381
PC3 (Energy-arousal)	0.798	0.435	0.494

Table 1: The mean correlation performance of the regression model at predicting the response PC from both acoustic and EEG features, acoustic features alone, and EEG features alone. All results are highly statistically significant ($p < 0.001$), but the model performs notably well when trained on both acoustic and EEG features.

3.2 RFE Analysis

The RFE selector's ranking of the 637 features determined which of the music and EEG feature types were the most important to the regression model. For the acoustic features, the key and tempo of the song performed best for the model. Similarly, the Hjorth parameter of mobility was the most vital EEG feature for training the model. Additional rankings are summarized below in Table 2.

ACOUSTIC FEATURES	EEG FEATURES
Major/Minor Key ($\mu=1.00$)	Hjorth Mobility ($\mu'=10.00$)
BPM/Tempo ($\mu=2.00$)	Energy ($\mu'=76.50$)
Chroma ($\mu=43.48$)	Standard Deviation ($\mu'=143.00$)
Spectral Centroid ($\mu=89.00$)	Mean ($\mu'=162.00$)
Spectral Rolloff ($\mu=124.71$)	Entropy ($\mu'=181.00$)
MFCC ($\mu=171.50$)	Power Spectrum ($\mu'=238.00$)
Spectral Flatness ($\mu=174.29$)	
Spectral Bandwidth ($\mu=250.00$)	
Spectral Contrast ($\mu=284.18$)	
ZCR ($\mu=307.00$)	
Tonnetz ($\mu=331.50$)	

Table 2: A ranking of the acoustic and EEG features. The acoustic features were ranked in the selector from 1 to 352. The mean feature score of each of the 11 acoustic feature types is notated by μ , in which the lower the μ value, the more important the feature type was to the regression model. Similarly, the EEG features were ranked in the selector from 1 to 285. For the EEG, μ' denotes the mean feature score of each of the 6 EEG feature types.

3.3 Demographic Analysis

Because the subjects of this study had a variety of age and gender differences and age and gender have been noted to convey differences in affective responses to emotions conveyed by music (Vieillard & Gilet 2013; Hunter et al. 2011), we tested if participant gender and age affected our model’s results. Therefore, predictions were calculated on an individual, per-participant level. Correlations between participant gender and ages and the predictions were then noted.

T-tests proved that the model’s ability to predict emotional responses was not influenced by gender or age (p=0.413 and p=0.278 respectively).

The strength of the subjects’ emotional responses to the music as taken from the Likert-scale was also tracked. On a scale from 0 to 4, with 4 being an extremely strong emotional response and 0 being a very weak emotional response, analyses found that female participants had an average emotional response strength of 2.48 while males’ strengths were 1.99. This difference indicates that the female participants were 24.2% more emotionally expressive towards the music than the males (p<0.001). However, in comparing the emotional response strengths for ages 18-66, no statistically significant differences were found (p=0.250).

3.4 Pop Songs (Acoustic Test) For Valence

The need for predicting music-induced emotions at an individual level is vital to music therapy and prescribing people music to improve mood and mental health. Because this model achieved such a relatively high correlation between the predicted emotions and actual responses, we tested if the model could benefit music therapy by predicting which songs would elicit the most positive emotional response in the participants. Although the combination of EEG and acoustic features performed the best, the acoustic features alone yielded a relatively high correlation in comparison to previous research (Daly et al. 2015; Song & Dixon 2015). Thus, the acoustic features were tested to see if they could be used to predict responses to

pieces of music not previously listened to by the participants.

As stated in 2.1, the first PC, valence, describes a musical stimuli’s positiveness (McConnell & Shore 2010). For example, music with higher valence induces more happy emotions than clips with low valence.

Given that most people listen to pop music over classical music and because music therapy most likely requires music longer than 15 seconds, the model was tested on a random set of longer, three-minute pop songs, as opposed to the original, short classical music excerpts. The model was trained on the acoustic features extracted from the classical music in 2.3 and tested on the held-out, acoustic features extracted from the pop songs. The model was then isolated for valence and ordered the songs from largest to smallest valence, thus indicating which songs would elicit the most cheerful emotional responses in the 31 participants of this study.

The model’s performance in predicting participant’s valence responses to pop songs they’ve never heard, based on their emotional responses to short classical music clips is summed up in Table 3.

Song Name	Artist	Valence Rating	Valence Order
Dancing Queen	ABBA	1.000	High Valence
Jump	Van Halen	0.985	
Stayin' Alive	Bee Gees	0.973	
Good Life	OneRepublic	0.965	
Happy	Pharrell Williams	0.745	
Uptown Girl	Billy Joel	0.606	
Yesterday	Beatles	0.580	
Say Something	A Great Big World	0.054	
Unchained Melody	Righteous Brothers	0.003	
Tears in Heaven	Eric Clapton	0.000	Lower Valence

Table 3: The model was trained on acoustic features extracted from short classical music clips and ordered a random selection of 10 pop songs from highest to lowest. As valence rating increases, the song’s level of induced-happiness increases.

IV. DISCUSSIONS

Although music therapy is an effective treatment for poor mental health disorders and depression (Maratos et al. 2008), it does not currently employ a systematic method to predict emotional responses to music on an individual level. Predicting

human emotional responses to music at high accuracy is a challenging problem because factors such as a person's age, gender, mood, and memories may affect how people emotionally respond to music. This challenge is heightened by combining the acoustic features of the music with the non-stationariness and noisiness of EEG.

Results from this study show that by combining EEG-derived features with acoustic features, emotional responses to music can be predicted at a significantly higher accuracy. This suggests that emotional responses to music are not just based on the musical properties of the music, but also the listener's idiosyncrasies and internal processes.

By training a regression model with the most performing EEG and acoustic features, music-induced emotions can be predicted at a higher accuracy than previously reported. For example, this experiments' outcomes were compared to [Daly et al. \(2015\)](#), which used the same dataset. By processing the EEG and audio differently and extracting unique and performing features only, our model achieved much better correlations. For example, for valence-arousal, they achieved a correlation of $24.3\% \pm 0.5\%$, while we achieved a correlation of 77.4% . They achieved a mean correlation of $15.8\% \pm 0.6\%$ for energy-arousal, while our model reached a correlation of 79.1% . Finally, for tension-arousal, while they only achieved an accuracy of $10.2\% \pm 0.5\%$, we met a mean correlation of 79.8% . This represents a 370% increase in accuracy over [Daly et al. \(2015\)](#).

Other researchers have analyzed emotions elicited via the DEAP dataset, using EEG to predict music video-induced emotions ([Nascimben et al. 2019](#); [Kumar et al. 2016](#)). However, these experiments extracted features solely from the EEG data and used music videos as opposed to just music as stimuli. Ultimately, [Nascimben et al. \(2019\)](#) achieved a cross-

validation accuracy of 65.4%. Additionally, ([Kumar et al. 2016](#)) used a single cross-validation run to achieve accuracies of 57.6% for valence-arousal and 62.0% for arousal. Our model is 32% more accurate.

Age and gender differences have been noted to affect emotional responses to music (for example, [Vieillard & Gilet \(2013\)](#) and [Hunter et al. \(2011\)](#)). While there were no differences in emotional responses across age, emotional responses across gender varied heavily, thus supporting [Hunter et al. \(2011\)](#). The strength of female responses to music was 24.2% more intense than those of males. This indicates that females are more emotionally expressive to music than men. That being said, our model's prediction accuracy for music-induced emotions was still consistent across gender and age.

Results from testing our model on a variety of pop songs show that this model can successfully be used in a clinical setting to improve music therapy techniques. For example, by giving people small selections of songs and recording their emotional responses to the clips on eight axes, our model can prescribe them a list of songs targeting specific emotions, e.g. happiness. Ideally, these treatments would also use patients' EEG as a parameter, but given that EEG may be expensive and infeasible, it is not necessary (as demonstrated in [3.4](#)).

Ultimately, the model created in this experiment can predict emotional responses to music on an individual level at significantly higher correlations than previously recorded. Our model has the potential to work as a music system to prescribe songs to patients to induce happier and more pleasant emotional states and advance music therapy as a treatment for mental health disorders and the emerging field of brain-computer music interfaces. Future work will strive to use these findings to create machine-generated music with acoustic properties that induce certain emotional states.

ACKNOWLEDGMENTS

The author would like to thank Tyler Giallanza for his support and contributions.

REFERENCES

1. Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains. *ISRN Neuroscience*, 2014, 1-7. doi:10.1155/2014/730218
2. Anggraeni, D., Sanjaya, W. S., Nurasyidiek, M. Y., & Munawwaroh, M. (2018). The implementation of speech recognition USING Mel-Frequency Cepstrum COEFFICIENTS (mfcc) and support vector Machine (SVM) method based on Python to CONTROL robot arm. *IOP Conference Series: Materials Science and Engineering*, 288, 012042. doi:10.1088/1757-899x/288/1/012042
3. Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, 58(1), 373-403. doi:10.1146/annurev.psych.58.110405.085709
4. Bauer, A. R., Kreutz, G., & Herrmann, C. S. (2014). Individual musical Tempo preference correlates with EEG beta rhythm. *Psychophysiology*, 52(4), 600-604. doi:10.1111/psyp.12375
5. Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., & Tervaniemi, M. (2011). A functional mri study of happy and sad emotions in music with and without lyrics. *Frontiers in Psychology*, 2. doi:10.3389/fpsyg.2011.00308
6. Brouwer, A., Van Wouwe, N., Mühl, C., Van Erp, J., & Toet, A. (2013). Perceiving blocks of emotional pictures and sounds: Effects on physiological variables. *Frontiers in Human Neuroscience*, 7. doi:10.3389/fnhum.2013.00295
7. 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. <https://pubmed.ncbi.nlm.nih.gov/1478135/>
8. Craig, D. G. (2005). An exploratory study of physiological changes during “chills” induced by music. *Musicae Scientiae*, 9(2), 273-287. doi:10.1177/102986490500900207
9. Daly, I., Pichiorri, F., Faller, J., Kaiser, V., Kreilinger, A., Scherer, R., & Muller-Putz, G. (2012). What does Clean EEG look like? *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. doi:10.1109/embc.2012.6346834
10. Daly, I., Williams, D., Hallowell, J., Hwang, F., Kirke, A., Malik, A., Weaver, J., Miranda, E., Nasuto, S. J. Music-induced emotions can be predicted from a combination of brain activity and acoustic features. *Brain Cogn.* 2015 Dec; 101:1-11. doi: 10.1016/j.bandc.2015.08.003.
11. Daly, I., Williams, D., Hwang, F., Kirke, A., Miranda, E. R., & Nasuto, S. J. (2019). Electroencephalography reflects the activity of subcortical brain regions during APPROACH-WITHDRAWAL behaviour while listening to music. *Scientific Reports*, 9(1). doi:10.1038/s41598-019-45105-2
12. Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A., Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, 19(7), 573-576. doi:10.1016/j.cub.2009.02.058
13. 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. doi:10.1177/0305735610362821
14. 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*, 61(1), 57-69. doi:10.1016/j.ijpsycho.2005.10.025.
15. Gabrielsson, A., & Lindström, E. (2001). *The influence of musical structure on emotional expression*. In P. N. Juslin & J. A. Sloboda (Eds.), *Series in affective science. Music and emotion: Theory and research* (p. 223-248). Oxford University Press.
16. Gosselin, N., Peretz, I., Hasboun, D., Baulac, M., & Samson, S. (2011). Impaired recognition of musical emotions and facial expressions following anteromedial temporal lobe excision. *Cortex*, 47(9), 1116-1125. doi:10.1016/j.cortex.2011.05.012
17. Hassani, M., & Karami, M. (2015). Noise estimation in electroencephalogram signal by using volterra series coefficients. *Journal of Medical Signals & Sensors*, 5(3), 192. doi:10.4103/2228-7477.161495
18. 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, Creativity, and the Arts*, 4(1), 47-56. <https://doi.org/10.1037/a0016873>

19. Hunter, P. G., Glenn Schellenberg, E., & Stalinski, S. M. (2011). Liking and identifying emotionally expressive music: Age and gender differences. *Journal of Experimental Child Psychology*, 110(1), 80-93. doi:10.1016/j.jecp.2011.04.001
20. Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23(4), 319-330. doi:10.1525/mp.2006.23.4.319
21. Ilie, G., & Thompson, W. F. (2011). Experiential and cognitive changes following seven minutes exposure to music and speech. *Music Perception*, 28(3), 247-264. doi:10.1525/mp.2011.28.3.247
22. Jang, D., Jin, M., & Yoo, C. D. (2008). Music genre classification using novel features and a weighted voting method. *2008 IEEE International Conference on Multimedia and Expo*. doi:10.1109/icme.2008.4607700
23. Jiang, X., Bian, G., & Tian, Z. (2019). Removal of artifacts from eeg signals: A review. *Sensors*, 19(5), 987. doi:10.3390/s19050987
24. Khalfa, S., Isabelle, P., Jean-Pierre, B., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, 328(2), 145-149. doi:10.1016/s0304-3940(02)00462-7
25. Kim, J., & Andre, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12), 2067-2083. doi:10.1109/tpami.2008.26
26. Koelsch, S., & Jäncke, L. (2015). Music and the heart. *European Heart Journal*, 36(44), 3043-3049. doi:10.1093/eurheartj/ehv430
27. Koelstra, S., Muhl, C., Soleymani, M., Jong-Seok Lee, Yazdani, A., Ebrahimi, T., . . . Patras, I. (2012). DEAP: A database for emotion Analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3(1), 18-31. doi:10.1109/t-affc.2011.15
28. Kortelainen, J., & Seppanen, T. (2013). EEG-based recognition of video-induced emotions: Selecting Subject-independent feature set. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. doi:10.1109/embc.2013.6610493
29. Kreutz, G., Ott, U., Teichmann, D., Osawa, P., & Vaitl, D. (2007). Using music to induce emotions: Influences of musical preference and absorption. *Psychology of Music*, 36(1), 101-126. doi:10.1177/0305735607082623
30. N. Kumar, K. Khaund, and S. M. Hazarika, "Bispectral Analysis of EEG for Emotion Recognition," *Procedia Computer Science*, vol. 84, pp. 31–35, 2016.
31. Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*, 81(4), 684-696. doi:10.1037/0022-3514.81.4.684
32. LeBlanc, A., Chang Jin, Y., Stamou, L., & McCrary, J. (1999). Effect of Age, Country, and Gender on Music Listening Preferences. *No. 141, The 17th International Society for Music Education: ISME Research Seminar*, 72-76.
33. Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y., & Hu, B. (2018). Exploring EEG features In Cross-Subject emotion recognition. *Frontiers in Neuroscience*, 12. doi:10.3389/fnins.2018.00162
34. Lundqvist, L., Carlsson, F., Hilmersson, P., & Juslin, P. N. (2008). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music*, 37(1), 61-90. doi:10.1177/0305735607086048
35. Maratos, A., S., Gold, C., Wang, X., & Crawford, M. J. (2008). Music therapy for depression. *The Cochrane Database of Systematic Reviews* (1), CD004517. doi: 10.1002/14651858.CD004517.pub2
36. McConnell, M. M., & Shore, D. I. (2010). Upbeat and happy: Arousal as an important factor in studying attention. *Cognition and Emotion*, 25(7), 1184-1195. doi:10.1080/02699931.2010.524396
37. McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J., & Gross, J. J. (2008). Gender differences in emotion regulation: An fmri study of cognitive reappraisal. *Group Processes & Intergroup Relations*, 11(2), 143-162. doi:10.1177/1368430207088035
38. Merrill, J., Omigie, D., & Wald-Fuhrmann, M. (2020). Locus of emotion influences psychophysiological reactions to music. *PLOS ONE*, 15(8). doi:10.1371/journal.pone.0237641

39. Milnik, V.. (2006). Instruction of electrode placement to the international 10-20-system. *Neurophysiologie-Labor*. 28. 113-143.
40. Nakakoga, S., Higashi, H., Muramatsu, J., Nakauchi, S., & Minami, T. (2020). Asymmetrical characteristics of emotional responses to pictures and sounds: Evidence from pupillometry. *PLOS ONE*, 15(4). doi:10.1371/journal.pone.0230775
41. Nascimben, M., Zoega Ramsay, T., & Bruni, L. E. (2019). User-Independent classification of emotions in a Mixed AROUSAL-VALENCE Model. *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*. doi:10.1109/bibe.2019.00086
42. Ramirez, R., Planas, J., Escude, N., Mercade, J., & Farriols, C. (2018). EEG-Based analysis of the emotional effect of music therapy on palliative Care cancer patients. *Frontiers in Psychology*, 9. doi:10.3389/fpsyg.2018.00254
43. Schaefer, H. (2017). Music-evoked emotions—current studies. *Frontiers in Neuroscience*, 11. doi:10.3389/fnins.2017.00600
44. Schimmack, U. and Grob, A. (2000), Dimensional models of core affect: a quantitative comparison by means of structural equation modeling. *Eur. J. Pers.*, 14: 325-345. [https://doi.org/10.1002/1099-0984\(200007/08\)14:4](https://doi.org/10.1002/1099-0984(200007/08)14:4)
45. 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. doi:10.1080/02699930126048
46. Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, 21(4), 561-585. doi:10.1525/mp.2004.21.4.561
47. Song, Y., & Dixon, S. (2015). How Well Can A Music Emotion Recognition System Predict the Emotional Responses of Participants? *Queen Mary University of London EECs*. doi:<https://www.eecs.qmul.ac.uk/~simond/pub/2015/SongDixon-SMC2015-EmotionPrediction.pdf>
48. Stevens, S. S., Volkman, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3), 185-190. doi:10.1121/1.1915893
49. Tamplin, J., & Baker, F. (2006). *Music Therapy Methods in Neurorehabilitation: A Clinician's Manual*. <https://njmt.w.uib.no/2007/08/20/music-therapy-methods-in-neurorehabilitation-a-clinicians-manual/>
50. Theimer, W., Vatulkin, I., & Eronen, A. (2008). Definitions of Audio Features for Music Content Description. *Faculty of Computer Science Algorithm Engineering (LS 11)*, 18.
51. van Tricht, M. J., Smeding, H. M., Speelman, J. D., & Schmand, B. A. (2010). Impaired emotion recognition in music in Parkinson's disease. *Brain and Cognition*, 74(1), 58-65. doi:10.1016/j.bandc.2010.06.005
52. Vieillard, S., & Gilet, A. (2013). Age-related differences in affective responses to and memory for emotions conveyed by music: A cross-sectional study. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00711
53. Vyas, Garima & Kumari, Barkha. (2013). SPEAKER RECOGNITION SYSTEM BASED ON MFCC AND DCT. *International Journal of engineering and advanced technology*. 2. 167-169.
54. Qazi, E., Hussain, M., Aboalsamh, H., Abdul, W., Bamatraf, S., & Ullah, I. (2016). An intelligent system to classify epileptic and NON-EPILEPTIC Eeg signals. *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. doi:10.1109/sitis.2016