

BRAIN-ACTIVITY-DRIVEN REAL-TIME MUSIC EMOTIVE CONTROL

Sergio Giraldo, Rafael Ramirez

Music Technology Group
Universitat Pompeu Fabra, Barcelona, Spain
sergio.giraldo@upf.edu

Abstract

Active music listening has emerged as a study field that aims to enable listeners to interactively control music. Most of active music listening systems aim to control music aspects such as playback, equalization, browsing, and retrieval, but few of them aim to control expressive aspects of music to convey emotions. In this study our aim is to enrich the music listening experience by allowing listeners to control expressive parameters in music performances using their perceived emotional state, as detected from their brain activity. We obtain electroencephalogram (EEG) data using a low-cost EEG device and then map this information into a coordinate in the emotional arousal-valence plane. The resulting coordinate is used to apply expressive transformations to music performances in real time by tuning different performance parameters in the *KTH Director Musices* rule system. Preliminary results show that the emotional state of a person can be used to trigger meaningful expressive music performance transformations.

Keywords: EEG, emotion detection, expressive music performance

1. Introduction

In recent years, active music listening has emerged as a study field that aims to enable listeners to interactively control music. While most of the work in this area has focused on control music aspects such as playback, equalization, browsing and retrieval, there have been few attempts to controlling expressive aspects of music performance.

On the other hand, electroencephalogram (EEG) systems provide useful information about human brain activity and are becoming increasingly available outside the medical domain. Similarly to the information provided by other physiological sensors, Brain-Computer Interfaces (BCI) information can be used as a source for interpreting a person's emotions and intentions.

In this paper we present an approach to enrich the music listening experience by allowing listeners to control expressive parameters in

music performances using their perceived emotional state, as detected by a brain-computer interface. We obtain brain activity data using a low-cost EEG device and map this information into a coordinate in the emotional arousal-valence plane. The resulting coordinate is used to apply expressive transformations to music performances in real time by tuning different performance parameters in the *KTH Director Musices* rule system (Friberg, 2006).

2. Background

The study of users' interaction with multimedia computer systems has increased in recent years. Regarding music, Goto (Goto, 2007) classifies systems based on which actions a listener is able to control. He classifies music systems into playback, touch-up (small changes

over audio signal, e.g. equalization), retrieval, and browsing. A related research line is the development of systems for automatic expressive accompaniment capable of following the soloist performance expression and/or intention in a real-time basis. Examples of such systems are the ones proposed by Cont et al. (Cont, 2012) and Hidaka et al. (Hidaka, 1995). Both propose systems able to follow the intention of the soloist based on the extraction of intention parameters (excitement, tension, emphasis on chord, chord substitution, and theme reprise). However, none of the above mentioned systems measure the listener/soloist intention/emotion directly from brain activity.

In this paper we propose a system, which allows listeners to control expressive parameters in music performances using their perceived emotional state, as detected from their brain activity. From the listener's EEG data we compute emotional descriptors (i.e. arousal and valence levels), which trigger expressive transformations to music performances in real time. The proposed system is divided in two parts: a real-time system able to detect listeners' emotional state from their EEG data, and a real-time expressive music performance system capable of adapting the expressive parameters of music based on the detected listeners' emotion.

2.1. Emotion detection

Emotion detection studies have explored methods using voice and facial expression information (K. Takahashi, 2004). Other approaches have used skin conductance, heart rate, and pupil dilation (Parala et.al, 2000). However, the quality and availability of brain computer interfaces has increased in recent years, making easier to study emotion using brain activity information. Different methods have been proposed to recognize emotions from EEG signals, e.g. (Chopin, 2000; Takahashi, 2004; Lin, 2010), training classifiers and applying different machine learning techniques and methods. Ramirez and Vamvakuosis (Ramirez, 2012) propose a method based on mapping EEG activity into the bi-dimensional arousal/valence plane of emotions

(Eerola, 2010). By measuring the alpha and beta activity on the prefrontal lobe, they obtain indicators for both arousal and valence. The computed values may be used to classify emotions such as happiness, anger, sadness, and calm.

2.2. Active music listening

Interactive performance systems have been developed in order to make possible for a listener to control music based on the conductor-orchestra paradigm. This is the case of the work of Fabiani (Fabiani, 2011) who use gestures to control performance. Gesture parameters are mapped to performance parameters adapting the four levels of abstraction/complexity proposed by Camurry et al. (Camurry, 2001). This level of abstraction range from low level parameters (physical level) such as audio signal, to high level parameters (semantic descriptors) such as emotions. Thus, gesture analysis is done from low to high level parameters, whereas synthesis is done from high to low level parameters. The control of mid and low level parameters of the performance is carried out using the KTH rule system by Fidberg (Friberg, 2006)

2.3. Expressive music performance

The study of music performance investigates the deviations introduced to the score by a skilled musician in order to add expression and convey emotions. Part of this research consists in finding rules to model these performance modifications that musicians use. Such is the case of the KTH rule system for music performance, which consists of a set of about 30 rules that control different aspects of expressive performance. These set of rules are the result of research initiated by Sundberg (Sundberg, 1983; Friberg, 1991; Sundberg, 1993). The rules affect various parameters (timing, sound level, articulation) and may be used to generate expressive musical performances. The magnitude of each rule is controlled by a parameter "k". Different combinations of k parameters levels model different performance styles, stylistic conventions or emotional intention. The result is a symbolic

representation that may be used to control a synthesizer. A real-time based implementation of the KTH system is the pDM (Pure Data implementation of Director Musices Profram) by Friberg (Friberg, 2006). Friberg implements an arousal/valence space control, defining a set of k values for the emotion at each quadrant of the space. Seven rules plus overall tempo and sound level are combined in such a way that they clearly convey the intended expression of each quadrant based on the research by Bresin et al. (Bresin, 2000) and Juslin (Juslin, 2001). Intermediate values for "k" are interpolated when moving across the space.

3. Methodology

Our proposed approach to real-time EEG-based emotional expressive performance control is depicted in Fig. 1. First, we detect EEG activity using the Emotiv Epoch headset. We base the emotion detection on the approach by Ramirez and Vamvakousis (Ramirez, 2012). We measure the EEG signal using electrodes AF₃, AF₄, F₃, and F₄, which are located on the prefrontal cortex. We use these electrodes because it has been found that the prefrontal lobe regulates emotion and deals with conscious experience.

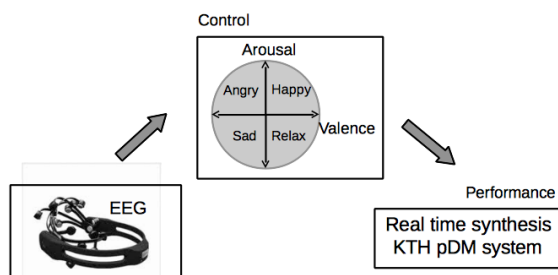


Figure 1. Theoretical frame work for expressive music control based on EEG arousal - valence detection.

We model emotion using the arousal-valence plane, a two dimensional emotion model which proposes that affective states arise from two neurological systems: arousal related to activation and deactivation, and valence related to pleasure and displeasure. In this paper we are interested in characterizing four different emotions: happiness, anger, relaxation, and sadness. As depicted in Figure 1,

each studied emotion belongs to a different quadrant in the arousal valence plane: happiness is characterized by high arousal and high valence, anger by high arousal and low valence, relaxation by low arousal and high valence, and finally sadness by low arousal and low valence.

3.1 Signal reprocessing

Alpha and Beta waves are the most often used frequency bands for emotion detection. Alpha waves are dominant in relaxed awake states of mind. Conversely Beta waves are used as an indicator of excited mind states. Thus, the first step in the signal preprocessing is to use a band pass filter in order to split up the signal in order to get the frequencies of interest, which are in the range of 8-12 Hz for alpha waves, and 12-30 Hz for beta waves.

After filtering the signal we calculate the power of each alpha and beta bands using the logarithmic power representation proposed by Aspiras & Asari (Aspiras et al., 2011). The power of each frequency band is computed by:

$$LP_f = 1 + \log\left(\frac{1}{N} \sum_{n=1}^N (x_{nf})^2\right)$$

Where x_{nf} is the magnitude of the frequency band f (alpha or beta), and N is the number of samples inside a certain window. Hence, we are computing the mean of the power of a group of N samples in a window and then compressing it by calculating the logarithm of the summation.

3.2 Arousal and valence calculation

After the band power calculation, the arousal value is computed from the beta/alpha ratio. Valence is calculated based on the asymmetric frontal activity hypothesis, where left frontal inactivation is linked to a negative emotion, whereas right frontal inactivation may be associated to positive emotions. Thus arousal and valence are calculated as follows:

$$arousal = \frac{b_{F3} + b_{F4}}{a_{F3} + a_{F4}}$$

$$valence = \frac{a_{F4}}{b_{F4}} - \frac{a_{F3}}{b_{F3}}$$

where b_{F3} , b_{F4} , a_{F3} and a_{F4} are respectively the beta and alpha logarithmic band power of electrodes F3 and F4.

The values obtained for arousal and valence are calculated using sliding windows over the signal in order to obtain a more smooth data. It is worth noting that there are not absolute levels for the maximum and the minimum values for both arousal and valence, as these values may differ from subject to subject, and also vary over time for the same subject. To overcome this problem we compute the mean of the last five seconds of a 20 second window and normalize the values by the maximum and minimum of these 20 sec window. This way we obtain values that range between minus one and one. We consider a window size of 4 seconds with 1 second hop size.

3.3 Synthesis

For synthesis we have used a real-time based implementation of the KTH group, pDM (Pure Data implementation of Director Musices Program) (Friberg, 2006). Thus, the coordinate on the arousal-valence space is mapped as an input for the pDM activity-valence space expressive control. In our implementation, this control is adapted in the pDM program, so the coordinates are rotated to fit the ones of the arousal valence space. Then the transformation of each of the seven expressive rules takes place by interpolating 11 expressive parameters between four extreme emotional expression values (Bressin and Friberg, 2000).

3.4 Experiments

Two types of experiments were performed: a first one listening while sitting down and motionless and the other listening while playing (improvising) with a musical instrument. In both the aim was to evaluate whether the in-

tended expression of the synthesized music corresponds to the emotional state of the user as characterized by his/her EEG signal. In both experiments subjects sat down in a comfortable chair facing two speakers. Subjects were asked to change their emotional state (from relaxed/sad to aroused/happy and vice versa). Each trial lasted 30 seconds with 10 seconds between trials. In experiment one the valence is set to a fixed value and the user tries to control the performance only by changing the arousal level. In experiment 2 the expression of the performance is dynamically changed between two extreme values (happy and sad), while the user is improvising playing a musical instrument. A 2-class classification task is performed for both experiments.

4. Results

The EEG signal and the corresponding calculated normalized arousal is shown in Figure 2. Vertical lines delimit the beginning and ending of each subtrial, and are labeled as "up" for high arousal and "down" for low arousal. The horizontal line represents the arousal average of each class segment. It can be seen how the calculated arousal corresponds to the intended emotion of the subject, and how the 2 classes can be separated by a horizontal threshold. However, further work should be done in order to obtain a smoother signal.

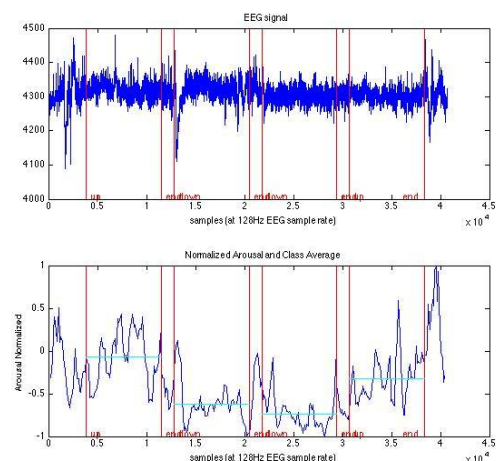


Figure 2. A subject's EEG signal (top) and calculated arousal (bottom). Vertical lines delimit each subtrial for high arousal (1st and 4th subtrials) and low arousal (2nd and 3rd subtrials). Horizontal line represents the average of each class segment.

Two classifiers, Linear Discriminant Analysis and Support Vector Machines, are evaluated to classify the intended emotions, using 10 cross fold validation. Initial results are obtained using the LDA and SVM implementations of the OpenVibe library (OpenVibe, 2010). Our aim was to quantify in which degree a classifier was able to separate the two intended emotions from the arousal/valence recorded data. For high-versus-low arousal classification we obtained a 77.23% for active listening without playing, and 65.86% for active listening when playing an instrument (improvising) along the synthesized expressive track, using SVM with radial basis kernel function. Results were obtained using 10-fold cross validation.

Initial results suggest that the EEG signals contain sufficient information to classify the expressive intention between happy and sad classes. However, the accuracy decreases, as expected, when playing an instrument. This may be due to the fact that the action of playing requires attention, thus, the alpha activity may remain low and beta may remain high

5. Conclusions

In this paper we have explored an approach to active music listening. We have implemented a system for controlling in real-time the expressive aspects of a musical piece, by means of the emotional state detected from the EEG signal of a user. We have performed experiments in two different settings: a first one where the user tries to control the performance only by changing the arousal level, and a second one where the performance is dynamically changed between two extreme values (happy and sad), while the user is improvising playing a musical instrument. We applied machine learning techniques (LDA and SVM) to perform a two class classification task between two emotional states (happy and sad). Initial results, in the first set where the subject was sitting still, suggest that EEG data contains sufficient information to distinguish between the two classes.

References

- Aspiras, T. H., & Asari, V. K. (2011). Log power representation of EEG spectral bands for the recognition of emotional states of mind. *2011 8th International Conference on Information, Communications & Signal Processing*, 1–5.
- Bresin, R., & Friberg, A. (2000). Emotional Coloring of Computer-Controlled Music Performances. *Computer Music Journal*, 24(4), 44–63.
- Camurri, A., Poli, G. De, Leman, M., & Volpe, G. (2001). A multi-layered conceptual framework for expressive gesture applications. *Proc. Intl MOSART Workshop*, Barcelona, Nov. 2001.
- Choppin, A. (2000). Eeg-based human interface for disabled individuals: Emotion expression with neural networks. *Master thesis*, Tokyo Institute of Technology, Yokohama, Japan
- Cont, A., & Echeveste, J. (2012). Correct Automatic Accompaniment Despite Machine Listening or Human Errors. In *Antescofo. International Computer Music Conference (ICMC)*. Ljubljana, Slovenia.
- 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.
- Fabiani, M. (2011). Interactive computer-aided expressive music performance. *PHD Thesis*, KTH School of Computer Science and Communication, Stockholm, Sweden. 2011
- Friberg, A. (1991). Generative Rules for Music Performance: A Formal Description of a Rule System. *Computer Music Journal*, 15(2).
- Friberg, A. (2006). pDM : An Expressive Sequencer with Real-Time Control of the KTH Music-Performance Rules. *Computer Music Journal*, 30(1), 37–48.
- Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology*, 2(2), 145–161.
- Goto, M. (2007). Active music listening interfaces based on signal processing. *The 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 2007, pp. IV–1441–1444).
- Hidaka, I., Goto, M., & Muraoka, Y. (1995). An Automatic Jazz Accompaniment System Reacting to Solo, *1995 International Computer Music Conference* (pp. 167–170).
- Juslin, P. 2001. Communicating Emotion in Music Performance: a Review and a Theoretical Framework. In *Juslin, P., and Sloboda, J., eds., Music and emotion: theory and research*. New York: Oxford University Press. 309–337.

Lin, Y., Wang, C., Jung, T., Member, S., Wu, T., Jeng, S., Duann, J., et al. (2010). EEG-Based Emotion Recognition in Music Listening. *IEEE Transactions on Biomedical Engineering*, 57(7), 1798–1806.

OpenViBE (2010). An Open-Source Software Platform to Design, Test, and Use Brain-Computer Interfaces in Real and Virtual Environments. *MIT Press Journal Presence* 19(1), 35–53.

Partala, T., Jokiniemi, M., & Surakka, V. (2000). Pupillary Responses To Emotionally Provocative Stimuli. *ETRA 00: 2000 Symposium on Eye Tracking Research & Applications* (pp. 123–129). New York, New York, USA: ACM Press.

Ramirez, R., & Vamvakousis, Z. (2012). Detecting Emotion from EEG Signals Using the Emotive Epos Device. *Brain Informatics Lecture Notes in Computer Science* (pp. 175–184). Springer.

Sundberg, J., Frydén, L., & Askenfelt, A. (1983). What tells you the player is musical? An analysis-by-synthesis study of music performance. In: J. Sundberg (Ed.), *Studies of Music Performance* (Vol. 39, pp. 61–75). Stockholm, Sweden: Publication issued by the Royal Swedish Academy of Music.

Sundberg, J., Askenfelt, A., & Frydén, L. (1983). Musical Performance. A synthesis-by-rule Approach. *Computer Music Journal*, 7, 37–43.

Sundberg, J. (1993). How Can Music be Expressive. *Speech Communication*, 13, 239–253.

Takahashi, K. (2004). Remarks on Emotion Recognition from Bio-Potential Signals. *2nd International Conference on Autonomous Robots and Agents*, 186–191.