An EEG-based Emotion-driven Music Control System

Macià Forteza Crespí

MASTER THESIS UPF/2013

Master in Sound and Music Computing

Master thesis supervisor:

Dr. Rafael Ramírez

Department of Information and Communication Technologies

Universitat Pompeu Fabra

Acknowledgments

I would like to thank Rafael Ramírez, Sergio Giraldo and Zackarias Vamvakousis for their help, patience and guidance. I want to express my gratitude to Fundació Catalunya Caixa for their financial support and to Xavier Serra for this opportunity. I also thank my family for their confidence in me and especially to *ma bibitte* for her support and encouragement.

Abstract

Brain-Computer Music Interfaces (BCMIs) aim to allow users to control music using their brain activity information. In this thesis the design and implementation of a BCMI for controlling the expressive content of musical pieces using emotions is presented. These emotions are obtained using encephalography (EEG) techniques. Human emotions can be characterized as a combination of arousal and valence values. However, the variability of these values across different subjects complicates the use of emotions for controlling music expression. Two experiments are presented to study the best approach for calculating arousal and valence value boundaries. The obtained results indicate that using images with emotional content in the process of calibrating the BCMI is less reliable than instructing the subjects to consciously modulate their excitation/relaxation state. Another conclusion is that the computed valence value is less reliable than the arousal value for controlling the BCMI. The impact of using both music and visual feedback in the BCMI is investigated and the main conclusion is that most of the users participating in the experiment are able to control better the BCMI when receiving only musical feedback compared to both music and visual feedback.

Table of contents

1. Introduction 1
1.1 Motivation 1
1.2 Research questions 1
1.3 Goals 2

2. State of the art 3
2.1 BCI 3
2.1.1 Basic Components of a BCI 3
2.1.2 Brain signals 4
2.1.3 Invasive and non-invasive BCIs 6
2.1.4 Recent non-invasive approaches 8
2.1.5 Neurofeedback 8
2.1.6 BCI-paradigms 9
2.1.7 BCI and emotion 9
2.2 BCMI 10
2.2.1 Direct Sonification 10
2.2.2 Musification 11
2.2.3 Control 11
2.3 Expressive music performance through emotions 12
2.3.1 Musical features and emotions 12
2.3.2 Programs to control music performance through emotions 14

3. Selected approach 16
3.1 Experiment 1 16
3.1.1 Materials 16
3.1.2 Image Selection 17
3.1.3 Experiment proceedings 17
3.1.3.1 Subjects 17
3.1.3.2 Test 17
3.1.3.3 Proceeding 19
3.1.3.4 Questionnaire 19
3.1.4 Feature extraction 19
3.1.4.1 Signal Acquisition 19
3.1.4.2 Spatial filtering 19
3.1.4.3 Frequency filtering 20
3.1.4.4 Logarithmic Power Representation 20
3.1.4.5 Arousal values 20
3.1.4.6 Valence values 21
3.1.5 Evaluation 21
3.1.6 Results 21
3.1.6.1 Questionnaire 21
3.1.6.2 EEG captured signals 22

3.1.7 Conclusions 23
3.2 Experiment 2 23
3.2.1 Image Selection 23
3.2.2 Experiment proceedings 24
3.2.2.1 Subjects 24
3.2.2.2 Test 24
3.2.2.3 Proceedings 24
3.2.3 Results 25
3.2.3.1 Questionnaire 25
3.2.3.2 Captured EEG signals 25
3.2.4 Conclusions 28
3.2.5 Comparison with other experiments 29
3.3 Emotion-based BCMI approach 31
3.3.1 Previous considerations 31
3.3.2 Materials 32
3.3.3 Emotion-based BCMI design 32
3.3.3.1 Signal Acquisition and Feature Extraction 32
3.3.3.2 BCI Application 32
3.3.3.3 Audio Module 34
3.3.4 Experiment Proceedings 34
3.3.5 Evaluation 36
3.3.6 Results 39
3.3.7 Conclusions 40
3.3.8 Repeating the experiment 41

4. Contributions and Future Work 42
4.1 Contributions 42
4.2 Future Work 42

References 43

Appendix 1 48
Appendix 2 49
Appendix 3 50
Appendix 4 55
Appendix 5 56
Appendix 6 62

List of figures

Figure 1. Basic Components of a BCI [23] 3

Figure 2. Brain signals in time domain [24]...5

Figure 3. The 10-20 electrode placement system [28] 7

Figure 4. Russel's Circumplex Model of Affect, also known as arousal-valence plane [35]........10

Figure 5. Mapping between musical features and emotions used in MOR2ART [21]................15

Figure 6. Static Arousal-Valence plane. The maximum and minumum values for arousal and valence are 200 and -200. 17

Figure 7. Selected IAPS Images. The circles define the selected images. 18

Figure 8. Names of the quadrants in the arousal-valence plane...18

Figure 9. Average of the questionnaire answers of all subjects. The blue circle represents the average value of the three images fitted in Q1 while the red cross represents the average value of the images fitted in Q2. The green star represents the images fitted in Q3 and the black plus symbol represents the Q4 images. 22

Figure 10. Image representation in the arousal-valence plane...23

Figure 11. Plot in the questionnaire meant to fill with a point where the image fitted well according to the subject...25

Figure 12. EEG averaged captured signals for Subject 1 in the first second after the stimulus was presented. 26

Figure 13. EEG averaged captured signals for Subject 2 in the three following seconds after the stimulus was presented. 27

Figure 14. Parts of the Emotion-based BCMI. 31

Figure 15. Visual feedback shown as a grid with an orange bar. The arrow is depicted in order to guide the subject in the calibration process. When up arrow appears, the subject tries to be as excited as possible. Conversely, when down arrow appears, the subject tries to be as relax as possible. 33

Figure 16. Visual feedback depicted in Training Mode...33

Figure 17. 2D arousal-valence space used in pDM to change the expressive music performance using the KTH rule system. 35

Figure 18. BCI Application first screen. The button ‘Calibration’ started the boundaries calculation process. ‘Training Mode’ button showed the grid with the orange bar and ‘Control Mode’ button allowed to control the musical excerpts. 36

Figure 19. High arousal period (in blue) and low arousal period (in red) captured from Subject 5’s brain signals during musical feedback scenario using polyphonic audio excerpt. Each period lasts 15 seconds. Blue flat line and red flat line represents the mean of high and low arousal periods, respectively. 37

Figure 20. High (blue) and low (read) arousal periods for each scenario. Flat lines are the mean values for each period. This plot corresponds to Subject 5’s brain signals. 37

Figure 21. High (blue) and low (red) arousal means for each scenario and the total high and low averages concerning Subject 5. The scenarios are Visual (V), Visual and Musical (V&M), No feedback (NoF) and Musical (M). ‘MIDI’ means a MIDI excerpt was played while ‘Audio’ means a polyphonic audio excerpt was played. 38

Figure 22. High (blue) and low (red) arousal periods and their mean values (flat lines) for all subjects during musical scenario using polyphonic audio excerpt. 38

Figure 23. High (blue) and low (red) mean values per subject and their average values (flat lines). 39

List of tables

Table 1. Relationship between musical features and emotions. Juslin and Laukka [15]. 12

Table 2. Main KTH rules for expressive performance [17]..13

Table 3. Mapping values between expressive musical features and emotions [20]. 14

Table 4. EEG captured signal values classification. Each column represents the accuracy for each subject while the last column represents the average of all subjects. 22

Table 5. Arousal classification using the first second after the stimuli. Each classification algorithm used 8 fold cross validation to train the model. 27

Table 6. Arousal classification using the three seconds after the stimuli. Each classification algorithm used 22 fold cross validation to train the model. 27

Table 7. Valence classification using the three seconds after the stimuli. Each classification algorithm used 22 fold cross validation to train the model. 28

Table 8. Valence classification using the first second after the stimuli. Each classification algorithm used 8 fold cross validation to train the model. 28

Table 9. Comparison between the experiments done in the state-of-the-art using IAPS and IADS stimuli and Experiment 2. 30

Table 10. High and Low arousal averages for each subject calculated from each individual scenario mean value. 39

Table 11. High and low arousal averages for each scenario calculated from each subject mean value...40

1. Introduction

1.1 Motivation

A Brain Computer Interface (BCI) is a device which captures the signals our brain emits. BCIs have been very important in medical research and several techniques have improved their use. For instance, BCIs can be used to explore the brain to look for physical defects or illness. Thanks to non-invasive techniques, BCIs have been used not only for medical issues, but also for spelling applications [1], virtual wheelchairs [2], [3], games [4] and driving performance enhancement [5]. Moreover, BCIs can be used to detect emotions [6], [7], [8], [9], [10].

A BCI involved in creating sound and music is referred to as a Brain Computer Music Interface (BCMI). BCMI research was started by Eduardo Reck Miranda and many approaches have been studied to create, compose or control music using brain signals [11], [12], [13], [14]. BCMIs can enhance the daily life of disabled people allowing them to create music through brain signals.

Music and Emotion studies the mapping between musical features and emotions [15], [16], [17], [18]. Which musical features represents one emotion or another? Which values of these musical features better express the desired emotion? How many musical features are necessary to cover one emotion? These questions have been answered during the last decades. Nowadays, there is an open discussion because all emotions cannot be represented by musical features. In fact, depending on the user, the same musical features can express different emotions.

The musical features used to express emotions are related to Expressive Music Performance. For example, changes in tempo, loudness or articulation make the listener feel certain emotions. There is a strong mapping between musical features and emotions that has been studied at KTH Royal Institute of Technology during the last decades. The most important musical features related to expressive performance are explained in the KTH rule system [17]. Selecting the appropriate rules, a computer system can change musical features to express emotions [19], [20], [21].

In this thesis, the relevant state-of-the-art of Music and Emotion and Expressive Music Performance are explained in the topic called Expressive Music Performance through Emotions.

1.2 Research questions

On one hand, we have BCIs capable of detecting human emotions. On the other hand, we can express emotions using *Music Expressive Performance*. Thus, the research questions are:

- Is it possible to control expression in music through emotions using BCIs?
- Can the user easily control an Emotion-based BCMI?

An Emotion-based BCMI is the program built to answer the first research question. It is called BCMI because it is an interface that creates music in real-time. It is called Emotion-based BCMI because the control over the application is done by means of emotions.

1.3 Goals

The goals of this thesis are:

- Build a BCMI which uses emotions as inputs to control expressive music performance.
- Evaluate the Emotion-based BCMI using objective and subjective evaluation. By objective evaluation we refer to giving an Emotion-based Task to the subject. The subject should be able to move a point throughout the arousal-valence plane at will. By subjective evaluation we refer to asking the user questions at the end of the session.

2. State of the art

The aim of this chapter is to introduce the previous research related to this thesis. First, we summarize the BCI research starting with a wide introduction and then focus on how BCIs are used to recognize and classify emotions. Secondly, we describe BCMIs emphasizing the BCMI classification proposed by Eduardo Reck Miranda [13]. Finally, we explore the expressive music performance topic focusing on emotions.

2.1 BCI

Brain Computer Interfaces (BCIs) began as a research topic by Jacques J. Vidal in 1973 [22]. Vidal demonstrated the feasibility of the communication between human encephalographic (EEG) signals and computers. This first step in brain-computer interaction opened the door to more accurate research and BCIs became a powerful tool to create many useful applications.

2.1.1 Basic Components of a BCI

The main goal of a BCI is to translate brain signals into messages for a computer [23]. This translation is done by following a set of steps: Signal Acquisition, Preprocessing, Feature Extraction and Classification (Figure 1). These steps create an output feedback, also called neurofeedback (see Neurofeedback) to better control the application.

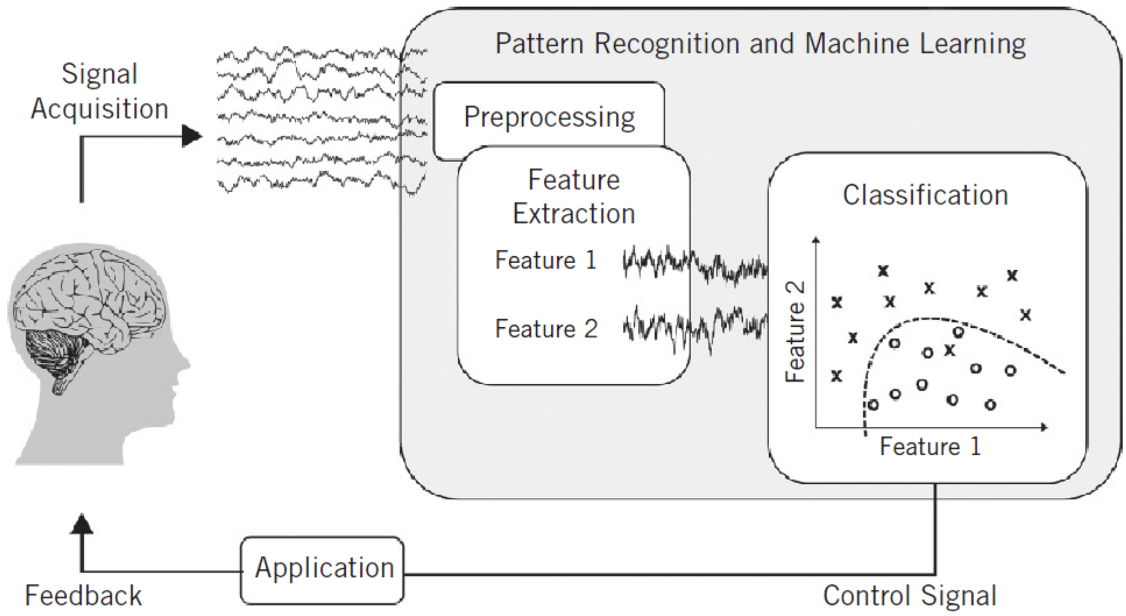


Figure 1. Basic Components of a BCI [23]

Signal Acquisition

To obtain the input brain signals, several techniques have been mentioned in BCI's research literature. The most commonly used technique is electroencephalography (EEG), which measures the difference in electric potentials between electrodes posed on the scalp. For more information, see Invasive and non-invasive BCIs.

Preprocessing

Brain signals contain significant levels of noise produced by the movement of facial muscle movement and, in the case of non-invasive BCIs, the skull and the scalp. Preprocessing techniques are thought to increase the signal-to-noise ratio (SNR) and erase noise and artifacts that disturb the input brain signals [23]. The most common preprocessing technique is spatial filtering by selecting the best electrodes of the BCI device.

Feature Extraction

The next step in BCI signal translation is the feature selection step. Normally, a band-pass filtering is applied to the signal such as *mu* (7-13 Hz), medium beta (13-30 Hz) or gamma (>30 Hz) rhythms [23]. For instance, regarding emotion classification (see BCI and emotion) arousal can be described as the division between beta rhythm (12-30 Hz) and alpha rhythm (8-12 Hz) of electrodes F3 and F4 [10]. The selection of the electrodes F3 and F4 corresponds to preprocessing step and the band-pass filtering (alpha and beta rhythms) and the division function correspond to feature extraction step. For a better understanding of electrode placement see Invasive and non-invasive BCIs and Figure 3.

Classification

The main goal of the classification step is to determine which output signal corresponds to each input signal. The input data can generate discrete outputs (e.g. LEFT-RIGHT-FORWARD) or continuous outputs (e.g. position or velocity of a prosthetic arm). The four most commonly used classifiers regarding machine learning techniques are Linear Discriminant Analysis (LDA), Regularized liner Discriminant Analysis (RDA), Quadratic Discriminant Analysis (QDA) and Support Vector Machines (SVM). SVM is the most frequently used in BCI research to classify input brain signals [23].

2.1.2 Brain signals

Depending on frequencies of the signals and states of consciousness, we can classify different brain signals. The information to explain different brain signals is extracted from [12] and [24]. Figure 2 depicts the brain signals.

Delta waves

The frequency of delta waves ranges between 0.5 Hz and 4 Hz. They are associated with sleeping. Additionally, delta waves are also useful to detect physical brain defects when they occur in a state of wakefulness.

Theta waves

Theta waves range from 4 Hz to 8 Hz. They are associated with emotional stress, such as disappointment, and also with deep relaxation or hypnosis.

Alpha waves

Alpha waves are associated with state of relaxation, especially when the eyes are closed. Their range is between 8 Hz and 13 Hz. Low levels of alpha waves also mean low levels of stress [25]. The *mu rhythm* is represented in this range and is associated with movement or intent to move. They are recorded over electrodes C3 and C4 which represent the motor cortex (see Invasive and non-invasive BCIs).

Beta waves

Beta waves are in the frequency range between 13 Hz and 30 Hz, but they are divided in two subgroups: low beta (13 Hz – 20 Hz) and medium beta (20 Hz – 30 Hz). They are associated with a high level of wakefulness. The more the beta wave's amplitude, the more concentration the subject has.

Gamma waves

Gamma waves, also known as high beta waves, have frequencies above 30 Hz. They represent a high level of consciousness, stress or anxiety.

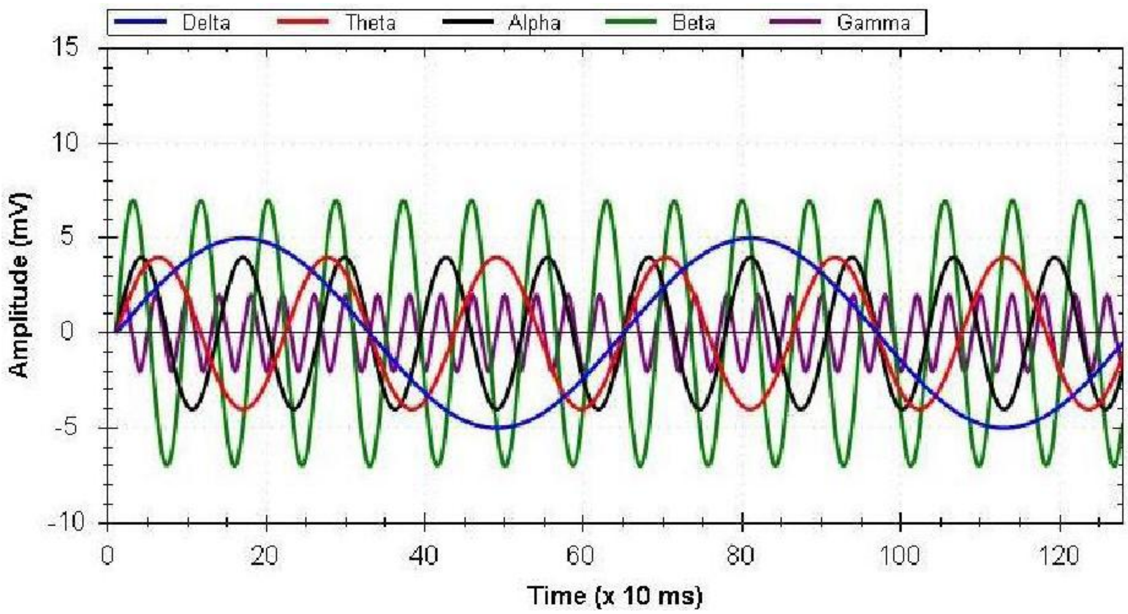


Figure 2. Brain signals in time domain [24]

2.1.3 Invasive and non-invasive BCIs

Invasive BCI techniques involve electrodes implanted intracranially. These techniques record the signals from a single area of the brain (e.g. motor cortex, auditory cortex) or from several areas at the same time. Invasive BCIs provide the best accuracy in temporal and spatial resolution but they have risks associated with surgical procedures [26].

On the other hand, non-invasive BCI techniques involves EEG signal recording without any surgical intervention. The non-invasive devices are based on electrodes placed on the surface of the head [23], [26]. This master thesis is based on a non-invasive BCI and this is the main reason to talk about non-invasive techniques that have been commonly used in BCI research in the last decade.

<u>Non-invasive BCI techniques</u>

Depending on the approach taken in capturing brain signals, non-invasive BCI techniques are divided into:

- **Electroencephalography** (EEG): The synapses of an ensemble of neurons generate electric potential that can be measured by an EEG. EEG signal processing has a high temporal resolution (on the order of milliseconds) but a bad spatial resolution (on the order of cm^3) due to the limited number of electrodes. Likewise, EEG devices are not expensive compared with the other non-invasive devices and are very portable [27].

  For a better understanding of EEG the reader should know about the 10-20 electrode placement system [12], [28]. This is an internationally recognized standard to place the EEG electrodes on the scalp. The '10' and the '20' refer to the distances (10% and 20%) between two adjacent electrodes regarding the head circumference. Figure 3 shows the 10-20 system and the distances between electrodes. The names of the electrodes are divided in two parts: letter and number. The letter represents the area of the brain where the electrode is placed: *Fp*, *F*, *C*, *T*, *P* and *O* represent *f*ronto*p*olar, *f*rontal, *c*entral, *t*emporal, *p*arietal and *o*ccipital lobe, respectively. Additionally, the letter 'z' represents the electrodes placed in the middle of the head. Even and odd numbers represent electrodes placed in the right and left hemisphere, respectively. Nz and Iz are two anatomical positions (naison and inion) but they are not electrodes. Nz and Iz are used as reference to place the electrodes correctly.

- **Magnetoencephalography** (MEG): Neural activity creates magnetic disturbance around the scalp. This magnetism can be measured by means of superconducting quantum interference device (SQUID) sensors. Approximately 300 SQUID sensors placed on the scalp in a non-portable device are needed to measure magnetic activity. MEG devices have high temporal resolution (milliseconds) and low spatial resolution (cm^3, slightly better than EEG devices). The most important factor is the price: since MEG devices cost between $2 and $3 million [27].

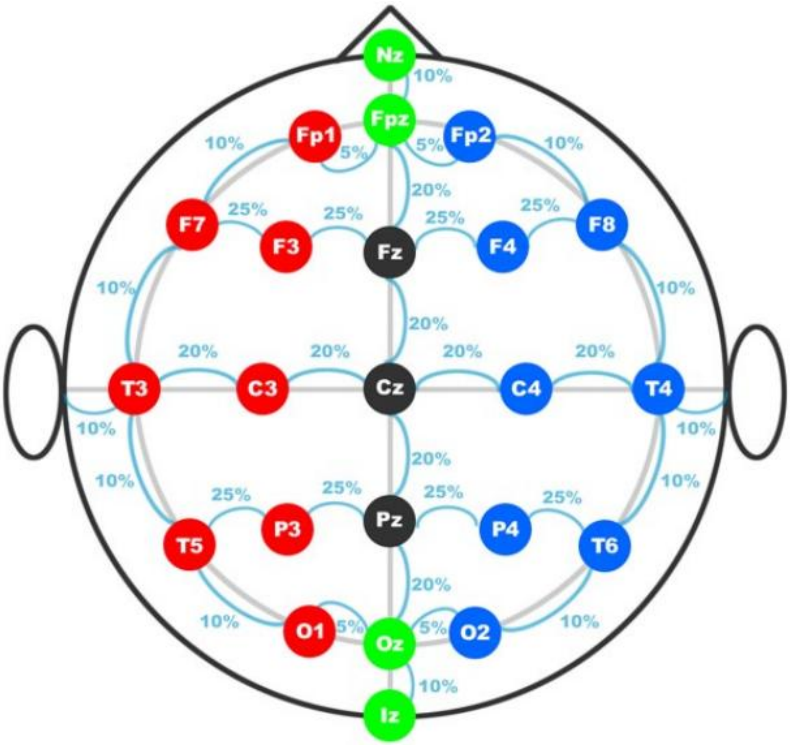


Figure 3. The 10-20 electrode placement system [28]

- **Functional Magnetic Resonance Imaging** (fMRI): fMRI signal processing is based on signal contrast detection in local blood oxygenation level (BOLD). Neural activity creates blood flow and fMRI devices capture these fluctuations. fMRI devices have low temporal resolution (1-2 seconds) and good spatial resolution (on the order of 64 mm^3). The portability and the price are again the major problems. They are non-portable and cost more than $1 million [27].

- **Near InfraRed Spectroscopy** (NIRS): NIRS devices capture changes in oxyhemoglobin (HbO_2) and deoxyhemoglobin (Hb) concentrations during neural activity by means of infrared spectrum light. The light penetrates the skull surface 1-3 cm deep and its intensity changes depending on the changes in NbO_2 and Hb. The NIRS sensors are called optodes. NIRS devices have medium temporal resolution (on the order of hundreds of milliseconds) and low spatial resolution (cm^3). They are portable and cheaper than fMRI and MEG devices [27].

- **Functional TransCranial Doppler sonography** (fTCD): When blood flows in the brain, it generates ultrasounds that can be detected by fTCD signal processing. fTCD devices are limited to the major vessels but they are growing in applicability. They have medium temporal resolution (tens of milliseconds) and low spatial resolution. Moreover, fTCD devices are portable and cheaper than MEG, fMRI and NIRS devices [27].

EEG, MEG and fMRI are the most commonly used devices for BCI signal processing in research. On the other hand, NIRS and fTCD are new techniques that are gaining popularity within BCI applications [27]. Furthermore, using EEG and fMRI together is another technique that benefits from EEG temporal resolution and fMRI spatial resolution.

This master thesis is focused on using an EEG device. Cost and usability are the main reasons for this choice.

2.1.4 Recent non-invasive approaches

BCIs can be used to recover movement control in paralyzed people, as described by Niels Birbaumer [29]. In his study, Birbaumer describes the use of invasive and non-invasive BCIs to 'break the silence' for people who cannot speak and/or move (e.g. people with amyotrophic lateral sclerosis or locked-in syndrome).

This 'silence' can be broken using a spelling application which uses imagined movement. 'Hex-o-spell' [30] was created by Benjamin Blankertz. This application allows the user to write using a non-invasive BCI which detects the intent to hand or foot movements. The user interface uses an arrow which moves clockwise by means of imagined hand movement and becomes higher when imaged foot movement is done. By means of the rotation and extension of the arrow, the subject selects different letters into different hexagons to write words. A physically disabled person needs only imagine a movement for writing everything.

Ferran Galán et al. used non-invasive BCIs to control a virtual wheelchair [2]. The goal was to avoid obstacles located in a virtual environment moving the virtual wheelchair by means of three different mental tasks. When the subject imagined a left hand movement, the virtual wheelchair moved leftwards. In order to move it rightwards, a word association task was applied. The third task, maintaining brain signals in a resting state, moved the wheelchair forwards. This approach demonstrates that the user can be trained to control his/her brain signals in order to achieve a goal.. Robert Leeb worked with tetraplegic subjects who used imaged hand movement, idling and imaged foot movement to control a virtual wheelchair in a virtual environment [3]. These two approaches demonstrate that sensory-motor activities can be done by means of BCIs.

BCIs can also be used for game applications. As seen before, a virtual environment can be used to control a virtual wheelchair. This virtual control paradigm becomes a game paradigm when game theory is applied to BCIs. Kenneth Oum et al. created "MindTactics", a BCI game which consisted of capturing flags in a virtual environment by means of concentration [4]. The interesting part of this approach is that the researchers could use this BCI game to study the brain signals behaviors.

To conclude this summary, we would like to mention Blankertz et al.'s review of non-medical applications using BCIs [5]. This review provides several examples of how to apply BCIs in non-medical issues. It explains the use of BCIs for measuring driving performance capability and how this approach can contribute to diminishing the number of driving accidents. In the same paper another approach explains the BCI game paradigm and how it can be used to study brain signals.

2.1.5 Neurofeedback

All the studies explained in Recent non-invasive approaches have two points in common: usage of BCI and the concept of neurofeedback.

Neurofeedback or EEG biofeedback 'usually refers to frequency-based biofeedback that uses an EEG to give clients information about their brainwaves and gradually and subtly teaches people how to alter their brainwave activity' as described by Deborah Stokes. In her study, 37 subjects were exposed to neurofeedback to control their migraines [31]. Pineda et al. [32] and Hwang et al. [33] used neurofeedback in their studies to control movement-based applications. They demonstrate that neurofeedback is a useful tool to enhance the communication between the subject and BCI devices.

In conclusion, it is important to take into account the effect of neurofeedback on subject's concentration [32]. Neurofeedback can disturb the main goal of the subject which is controlling the BCI; however, the level of disturbance depends on the subject himself/herself.

2.1.6 BCI-paradigms

It is important to distinguish between the two major BCI paradigms for a better understanding of this thesis.

- Cue-guided BCIs: The BCI signal processing is based on capturing the signals of the subject after presenting a set of stimuli to him/her. Normally, in cue-guided BCIs, these stimuli are pre-defined and the resulting signal is expected. This paradigm is off-line signal processing. The brain signals are analyzed after being captured [23].

- Self-paced BCIs: The BCI signals are directly controlled by the user. Due to neurofeedback, the user is able to change his/her mental state and sends the desired signals to the device to control it. The device should analyze the signal during capture time to automatically detect the signal changes. This is an on-line paradigm [23].

2.1.7 BCI and emotion

Emotions can be projected onto a two-dimensional plane where one of the dimensions is the valence (pleasure-displeasure dimension) and the other is the arousal (activation-deactivation) [34]. In literature this plane is called arousal-valence plane and the most famous model of emotion is the "Circumplex Model of Affect" proposed by James A. Russell [35]. On this plane, basic emotions like fear, anger, disgust, sadness and happiness can be plotted. Figure 4 depicts the model proposed by Russell.

Tuomas Eerola and Jonna K. Vuokoski demonstrated that Russell's model is the best model to represent emotions in a plane using sound stimuli [34]. In fact, several researchers used this plane in emotion classification studies. Danny Oude Bos used Fisher's Discriminant Analysis (FDA) to classify emotions in three classes (modality, valence and arousal) [6]. He used audio and visual stimuli data from IADS [36] and IAPS [37] datasets. Antoine Choppin proposed an emotion classification method using neural networks. The stimuli used in Choppin's study were also extracted from IAPS and IADS databases [7]. Kazuhiko Takahashi used Support Vector Machines (SVM) to classify joy, anger, sadness, fear and relaxation using audiovisual stimuli [8]. Yuan-Pin Lin et al. used SVM to classify joy, anger, sadness and pleasure using soundtracks from films that won Oscar's [9]. Eerola and Vuokoski, in their study to determinate the best model to classify emotion [34], used soundtracks excerpts stimuli to classify happiness, sadness, tenderness, fear and anger. Rafael Ramírez and Zackarias Vamvakousis classified emotions using three different classifiers (Linear Discriminant Analysis, LDA, and SVM) and the stimuli used were IADS dataset [10].

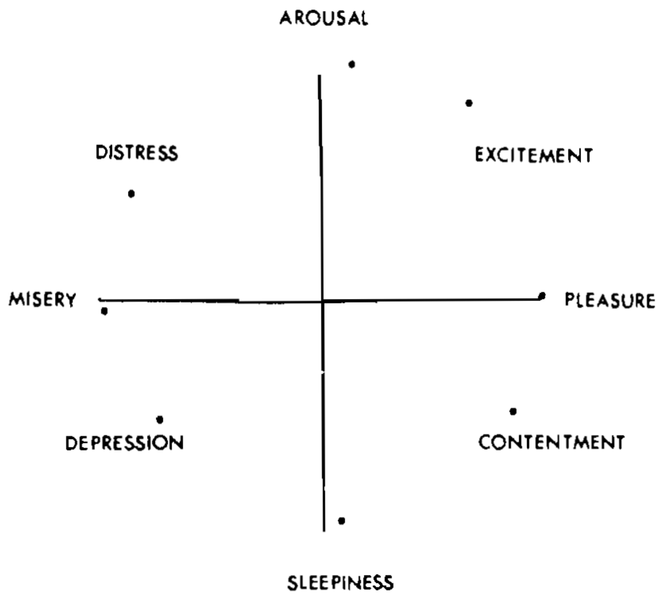


Figure 4. Russel's Circumplex Model of Affect, also known as arousal-valence plane [35]

These previous studies demonstrated that it is possible to classify emotions in an arousal-valence plane using brain signals.

2.2 BCMI

Brain Computer Music Interfaces (BCMIs) are 'BCI systems for musical applications paying attention in real-time music systems'. This concept was defined by Eduardo Reck Miranda and Andrew Brouse in their paper "Toward direct brain-computer music interfaces" (2005) [14].

BCMIs are classified in three groups, depending on the their approach in using the brain signals [13]: Direct Sonification, Musification and Control.

2.2.1 Direct Sonification

Direct Sonification is the technique that directly translates brain signals into sound [13]. The first piece that used brain waves was "Music for Solo Performer" by Alvin Lucier [11]. Lucier amplified the detected alpha waves and the resulting electrical signal was used to vibrate percussion instruments distributed around the performance space.

After Lucier many artists and researchers used brain waves to make music. Richard Teitelbaum used biological signals (heartbeat, breathing sounds and EEG waves) to control synthesizers [38]. In 1971, David Rosenboom developed a biofeedback collaborative application called "Ecology of the Skin" which used brain waves and heart signals that were translated into musical texture [39]. Adam Overton analyzed breaths, heartbeats and brain waves to create a noisy-chaotic musical texture [40].

In recent approaches, Thilo Hinterberger created POSER [41] and The Sensorium [42], two applications which used neurofeedback and *mu* rhythms to make what Hintergerber called 'brainmusic'. The brain signals were projected into loudspeakers and lightscapes so the user was receiving neurofeedback from his/her neurophysiologic rhythms.

2.2.2 Musification

Musification is the technique that 'translates EEG signals into music by a system that generates musical sequences based on the behavior of the EEG' [13]. Dan Wu et al. composed music based on sleep EEG signals [43]. EEG signals from 35 volunteers were recorded and by introducing musical knowledge they create understandable and non-stochastic melodies. In 2010, Eduardo Reck Miranda and his team developed a mapping to generate melodies assigning each EEG electrode with a musical note [12]. The higher amplitude of the electrode signal was the winner and the note related with this electrode was written in a score. Brahim Hamadicharef et al. [44] and Yee Chew and Eric Caspary [45] designed BCI-based musical composition applications. These applications had an interface with several buttons, one for each musical note or a grid with a set of notes. When the subject focused on a specific button, high amplitude in the signal appeared. The computer captured this high amplitude and wrote the musical note on a score.

2.2.3 Control

In control technique the main goal is to command a musical application by means of EEG signals [13]. In the first paper related to BCMI [14], Miranda and Brouse created a BCMI-piano, an engine that controls the composer-specific performance music played by means of alpha and beta waves. This BCMI-piano used generative music [46] to play one kind of composer-related music or another. For instance, if alpha waves were higher in amplitude than beta waves, the BCMI-piano played a Schumann-like performance. Conversely, if beta waves were higher in amplitude than alpha waves, it played a Beethoven-like performance. Again, in 2010, Miranda created a BCMI that controlled the volume of guitar and piano solos by means of alpha and beta waves [12]. In [13] the BCMI employed a user interface to control a set of notes pre-established by musical knowledge. It developed into an instrument when the user acquired more skill using the device.

BCMIs are applications that allow the subject to create music using only brain signals. It can be very helpful in enhancing the daily life of disabled people who cannot move their arms, legs or the entire body, such as people who suffer amyotrophic lateral sclerosis [13]. Vamvakousis and Ramírez created a Brain-Gazed controlled musical interface [47] that used brain signals to detect emotions and to map them onto an arousal-valence plane. Positive and negative valence values trigger major or minor sequences of chords, respectively. This application allows disabled people to compose their own music using their brain signals to create the background progression and their eyes to create the melodies.

In conclusion, music can be created without using legs, arms or mouth, as in classic instruments. BCMIs can become new instruments which offer a new way to play music: the brain signals.

2.3 Expressive music performance through emotions

Expressive music performance is the most important part for listeners when a musician performs. Without expression music makes no sense [48]. This expressive performance depends on variations in tempo, pitch, dynamics and timbre [49]. Other features such as variation in onset, energy or embellishment (several notes instead of only one note) are also important for expressive performance [50].

BCMIs could potentially use this relationship between expressive performance and emotions to create music using brain signals.

2.3.1 Musical features and emotions

This master thesis is focused on musical features related to emotions. Which music features are important to express emotions? Several researchers have been working to answer this question for years.

Patrick Juslin and Petri Laukka studied the relationship between vocal expression of emotions and musical expression of emotions. In [15] they presented a review of the main studies in these two fields. They conclude with a mapping between musical features and emotions (Table 1). Additionally, they demonstrated that 90% of emotions can be represented by means of tempo, dynamics and articulation. In [16] they studied the implications of music in everyday listening and demonstrated that music can express emotions.

Anders Fridberg, Roberto Bresin and Johan Sunders wrote a review of the KTH rule system [17]. The KTH rule system is an expressive model that explains which musical features cover Western classical music, jazz and popular music. During the last 20 years, the KTH Royal Institute of Technology has been studying these features and enhancing year by year their model using feedback from experts. Table 2 shows a review of the main KTH rules.

Other studies have focused their attention on a smaller number of features. Roberto Bresin and Anders Fridberg, two of the participants in the overview of the KTH rule system, studied seven features to express neutral, happy, scary, peaceful and sad emotions. These seven features were tempo, sound level, articulation, phrasing, register, timbre and attack speed [18]. They demonstrated that changing these seven features, the emotion of a song also changes independently of the initial emotion that the song tries to express. Marco Fabiani [19], used only three musical features to change the emotion of a melody: tempo, dynamics and articulation.

Thus, it has been demonstrated that emotions can be expressed using a reduced number of musical features. In this thesis we will focus on two musical features: tempo and dynamics.

Table 1. Relationship between musical features and emotions. Juslin and Laukka [15].

Emotion	Acoustic cues (vocal expression/music performance)
Anger	Fast speech rate/tempo, high voice intensity/sound level, much voice intensity/sound level

	variability, much high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and microstructural irregularity
Fear	Fast speech rate/tempo, low voice intensity/sound level (except in panic fear), much voice intensity/sound level variability, little high-frequency energy, high F0/pitch level, little F0/pitch variability, rising F0/pitch contour, and a lot of microstructural irregularity
Happiness	Fast speech rate/tempo, medium–high voice intensity/sound level, medium high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and very little microstructural regularity
Sadness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contour, slow voice onsets/tone attacks, and microstructural irregularity
Tenderness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contours, slow voice onsets/tone attacks, and microstructural regularity

Table 2. Main KTH rules for expressive performance [17].

Phrasing	
Phrase arch	Create arch-like tempo and sound level changes over phrases
Final ritardando	Apply a ritardando in the end of the piece
High loud	Increase sound level in proportion to pitch height
Micro-level timing	
Duration contrast	Shorten relatively short notes and lengthen relatively long notes
Faster uphill	Increase tempo in rising pitch sequences
Metrical patterns and grooves	
Double duration	Decrease duration ratio for two notes with a nominal value of 2:1
Inégales	Introduce long-short patterns for equal note values (swing)
Articulation	
Punctuation	Find short melodic fragments and mark them with a final micropause
Score legato/staccato	Articulate legato/staccato when marked in the score
Repetition articulation	Add articulation for repeated notes.
Overall articulation	Add articulation for all notes except very short ones
Tonal tension	
Melodic charge	Emphasize the melodic tension of notes relatively the current chord
Harmonic charge	Emphasize the harmonic tension of chords relatively the key
Chromatic charge	Emphasize regions of small pitch changes
Intonation	
High sharp	Stretch all intervals in proportion to size
Melodic intonation	Intonate according to melodic context
Harmonic intonation	Intonate according to harmonic context
Mixed intonation	Intonate using a combination of melodic and harmonic intonation
Ensemble timing	
Melodic sync	Synchronize using a new voice containing all relevant onsets
Ensemble swing	Introduce metrical timing patterns for the instruments in a jazz ensemble
Performance noise	
Noise control	Simulate inaccuracies in motor

2.3.2 Programs to control music performance through emotions

To finish this section we review three programs used in research to change expressive performance by means of changing emotions. Two of the three programs are based on MIDI (pDM, MOR2ART) and the third one is based on real audio (PerMORFer).

- pDM is a real-time application to play performance MIDI songs [20]. This performance uses the KTH rule system [17] to change the way in which the MIDI song is played. Not all the rules are applied due to the required computational work. The pDM is written in Pure Data [51] and the user interface allows the performance to be controlled through emotions thanks to an arousal-valence plane. When the user moves the point inside the arousal-valence plane using the mouse, the MIDI configuration changes depending on the position of the point. Table 3 shows the mapping between the four corner emotions (Happy, Tender, Sad and Angry) and rule values applied on the MIDI song.

Table 3. Mapping values between expressive musical features and emotions [20].

Rule	Happy	Tender	Sad	Angry
Phase Arch 5	1	1.5	3	-1
Phase Arch 6	1	1.5	3	-1
Final Ritard	0.5	0.5	0.5	0
Duration Contrast	1.5	0	-2	2.5
Punctuation	1.8	1.2	1	1.4
Repetition Articulation	2	1	0.8	1.5
Overall Articulation	2.5	0.7	0	1
Tempo Scaling	1.1	0.8	0.6	1.2
Sound Level Scaling	3	-4	-7	7

- MOR2ART [21] uses the arousal-valence plane to change the expressiveness of MIDI songs. With a wireless pointer, the user moves a point through the plane changing performance profiles such as timbre, tempo and loudness. Figure 5 depicts the arousal-valence plane and the associated musical features.

- Marco Fabiani created PerMORFer [19] capable of transforming monophonic excerpts in real-time according to some KTH rules [17]. PerMORfer manages three musical features to change the expressiveness of sounds using emotions: tempo, dynamics and articulation. The tempo is changed using time stretching algorithms, such as [52]; the dynamics are modified using spectral extension to increase or decrease sound level and brightness [53]; and articulation changes are achieved using staccato-legato transformations. An arousal-valence plane is the controller to automatically change these parameters.

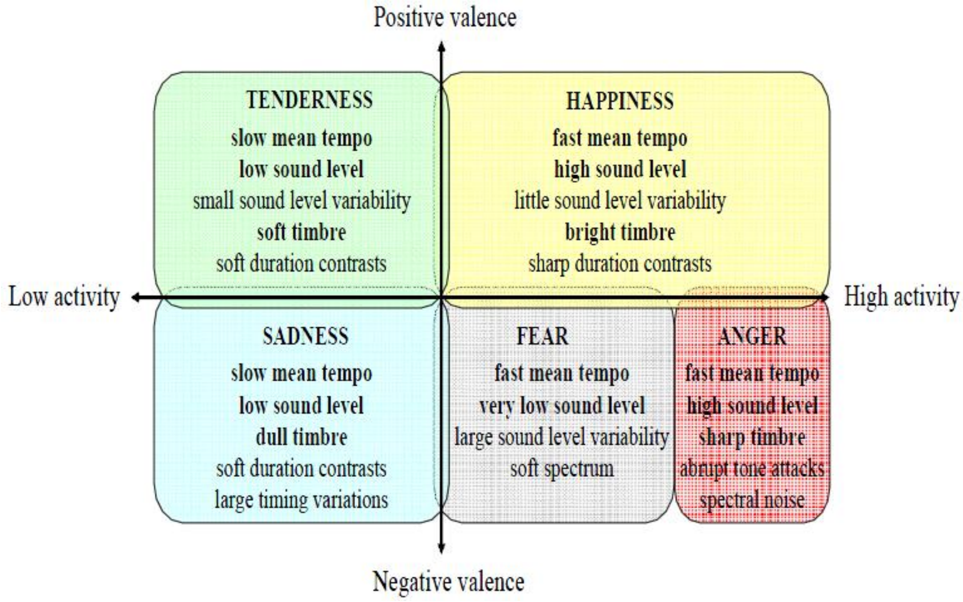


Figure 5. Mapping between musical features and emotions used in MOR2ART [21]

3. Selected approach

The aim of this section is to describe the followed steps for answering the research questions:

- Is it possible to control expression in music through emotions using BCIs?
- Can the user easily control an Emotion-based BCMI?

One should divide the first research question in two different parts: (1) Selecting the stimuli to evoke emotions and (2) Mapping the captured EEG signals in the arousal-valence plane. In (1) the main goal is to show the subject a set of stimuli to capture which are the highest and lower arousal and valence values. Once these values are captured, in (2) they can be mapped into the arousal-valence plane. For instance, the arousal-valence plane has its maximum arousal and valence values at 200 and -200; (0,0) is the point in the middle (Figure 6). A mapping or a concordance between the captured EEG signals and the values of the arousal-valence plane is needed because every subject has different EEG values. Imagine that the highest arousal value of one subject is 2.5 and the lower is 1.3. As a consequence, the mapping should determine that 2.5 is equal to 200 and 1.3 is -200. This mapping is called boundaries calculation.

In conclusion, the main problem to solve was to determine which the boundaries are for every subject so the first step was to calibrate the Emotion-based BCMI. Experiment 1 and 2 are referred to the calibration process and the Emotion-based BCMI approach describes how to build the BCMI application.

3.1 Experiment 1

The Experiment 1 is based on the experiments proceedings of [6] and [10]. The goal was to determine if images are enough stimuli for selecting the boundaries of arousal a valence. This approach tried to demonstrate if the selected images are defined properly to fit in the four quadrants of the arousal-valence plane and if the EEG captured values for each image will correlate with the supposed quadrant where the image is fitted.

3.1.1 Materials

In order to capture the brain signals, a 14-electrode neuroheadset called Emotiv EPOC[1] was used. This device follows the aforementioned 10-20 electrode placement and uses a bipolar measurement using two reference electrodes. The Emotiv EPOC device sends the brain signals to a PC via wireless.

After capturing brain signals, OpenViBE [54], an open-source software meant to create BCI applications, was used to calculate the arousal and valence values and show the images to the subject. Additionally, this software stored the brain signals into GDF files in order to recall them for evaluation purposes.

[1] http://www.emotiv.com/epoc/features.php

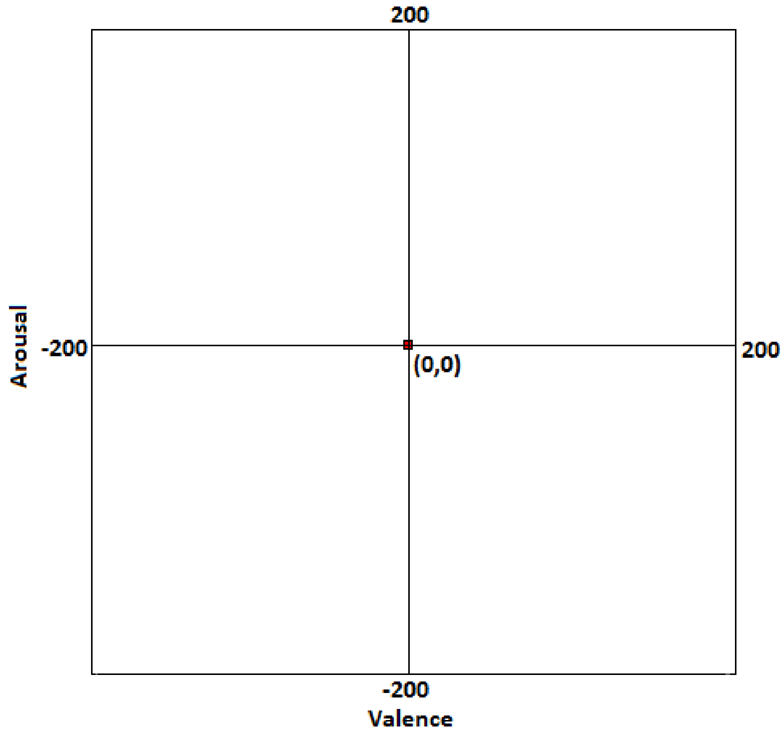


Figure 6. The Arousal-Valence plane. The maximum and minumum values for arousal and valence are 200 and -200.

3.1.2 Image Selection

Previous research used the International Affective Picture System (IAPS) to evoke emotion and to classify the arousal and valence EEG captured values [6], [7]. The challenge was to define images with enough impact and emotional charge but without injury. The images were selected by hand searching the suitable ones which better represents the four quadrants in the arousal-valence plane. Figure 7 depicts the selected images from IAPS dataset. Three images per quadrant were selected.

3.1.3 Experiment proceedings

3.1.3.1 Subjects

Ten subjects aged between 23 and 45 years old participated in the experiment. Nine of ten were males and the other one was female. All of them were members of the Music Technology Group at Universitat Pompeu Fabra.

3.1.3.2 Test

12 images were showed in front of the subject lasting 15 seconds each. A 6 seconds black image was shown in between two following images. Each image stimuli corresponded with a quadrant in the arousal-valence plane and in order to avoid emotional subject accommodation, these images were shown in the following order: Q1, Q2, Q3, Q4, Q2, Q3, Q4, Q1, Q3, Q4, Q1 and Q2. Letter ‘Q’ means quadrant. Figure 8 depicts the correspondence between the quadrant and its name in the arousal-valence plane.

During the test, the EEG device captured the brain signals and stored the values into a computer. Each transition between images was marked for the easier subsequent analysis. Additionally, the subject was asked to fill a questionnaire which consisted on determining the subjective arousal and valence values.

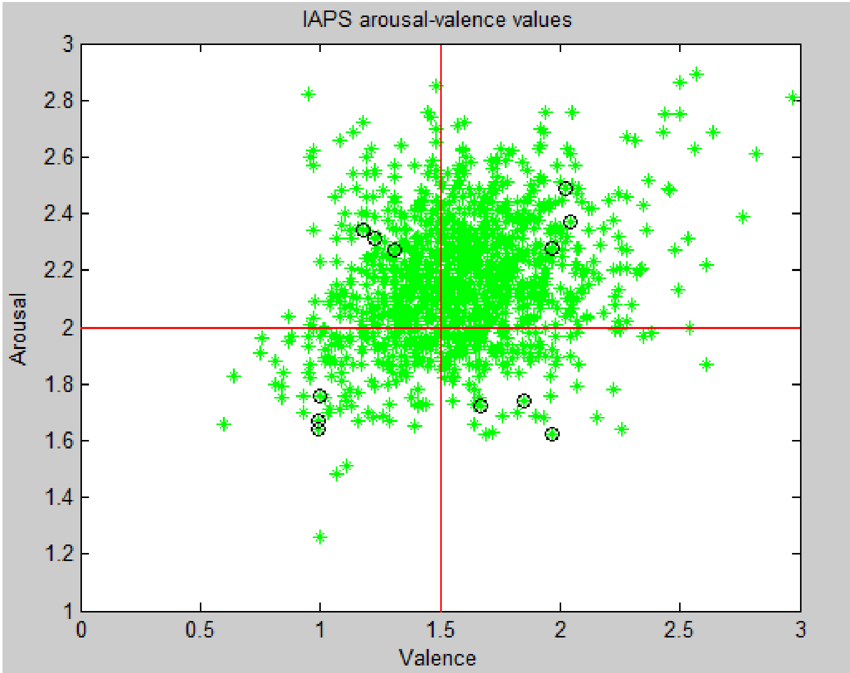


Figure 7. Selected IAPS Images. The circles define the selected images.

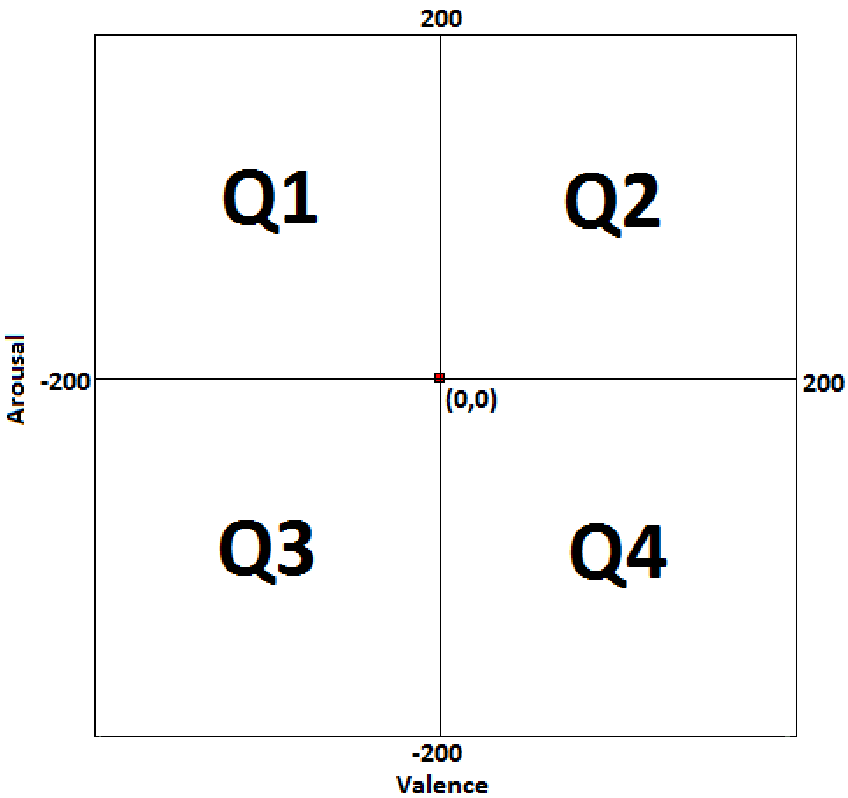


Figure 8. Names of the quadrants in the arousal-valence plane

3.1.3.3 Proceeding

The subject was informed to sit in front of a computer screen where the images were shown. The first step was to be sure that the subject was feeling comfortable for filling the questionnaire moving himself/herself as less as possible. The next step was to pose the Emotiv EPOC device on the scalp of the subject and to guarantee that all the electrodes were working properly. Additionally, the subject was instructed to blink and move as less as possible and to not speak during the test.

3.1.3.4 Questionnaire

Before beginning the experiment, the subject was informed to fill the questionnaire during the 6 seconds black image placed in between two following image stimuli. The questionnaire consisted in 12 rows each of them representing the 12 selected images. In each row three columns were depicted: (1) The number of the image, (2) arousal value and (3) valence value. Arousal value ranged from 1 to 5, being 1 the less and 5 the most, while valence value ranged also from 1 to 5 but the subject was meant to fill circles instead of selecting numbers. The questionnaire template can be shown in Appendix 1.

3.1.4 Feature extraction

Feature extraction means to calculate the valence and arousal values after capturing the EEG signals. In this experiment the selected approach to extract arousal and valence values was the used in [10].

3.1.4.1 Signal Acquisition

Emotiv EPOC Device was posed on the scalp of the subject and connected to the computer. Every 128 milliseconds one value per electrode was sent to the computer software and stored in a file.

After the signal acquisition, in order to calculate arousal and valence values, the signal was analyzed using a 0.5 seconds window size and a 0.25 seconds hop-size. As a consequence, four values per seconds were obtained.

3.1.4.2 Spatial filtering

The most common used electrodes to capture emotion are AF3, AF4, F3 and F4 (Figure 9). A spatial filtering to select these electrodes was applied to the captured signal.

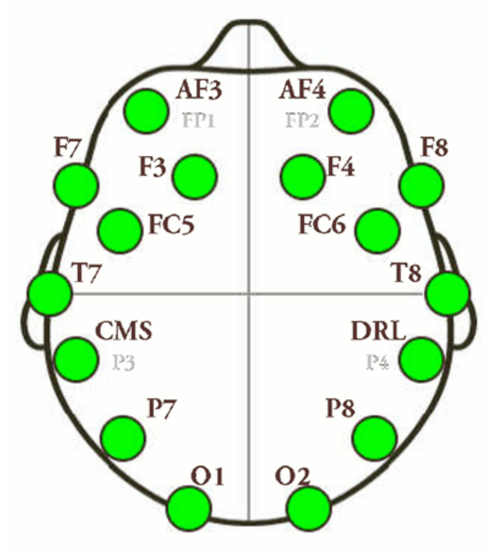


Figure 9. Scalp locations covered by Emotiv EPOC neuroheadset.

3.1.4.3 Frequency filtering

Alpha and Beta waves are necessary to calculate arousal and valence values. Alpha ranges from 8 Hz to 13 Hz and Beta from 13 Hz to 30 Hz. Two frequency filters per electrode (just the selected ones in the spatial filtering step) were applied.

3.1.4.4 Logarithmic Power Representation

Having applied the spatial and frequency filters the logarithmic power representation proposed by Theus H. Aspiras and Vijayan K. Asari [55] is computed by applying Eq1. This formula calculates the mean of the power of number of samples in a window and compressing it by calculating the logarithm of the summation. As a consequence, all the calculated values are positive.

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

3.1.4.5 Arousal values

Arousal is associated with an excited state of mind. In literature its values are captured using electrodes posed in the prefrontal cortex (AF3, AF4, F3 and F4) and calculating an alpha-beta ratio**.** Eq2 depicts the formula to calculate arousal values.

$$\frac{\beta AF3 + \beta AF4 + \beta F3 + \beta F4}{\alpha AF3 + \alpha AF4 + \alpha F3 + \alpha F4} \quad (2)$$

3.1.4.6 Valence values

Valence corresponds with a pleasure state of mind. The less the pleasure is, the smaller the valence value. In literature, one of the ways to calculate valence is comparing the activation levels between the two brain hemispheres. F3 and F4 represent the prefrontal cortex in left and right hemisphere, respectively. Alpha and Beta frequencies bands are used to estimate valence values (Eq3).

$$\frac{\alpha F4}{\beta F4} - \frac{\alpha F3}{\beta F3} \tag{3}$$

3.1.5 Evaluation

In order to evaluate if the images fitted in the four quadrants, the proposed method was divided into two evaluation approaches: subjective and objective.

The subjective evaluation was based on the questionnaire. A level of arousal and valence per image was obtained after answering the subject. The idea was to observe if these levels were correlated with the quadrants where the images were fitted.

The objective evaluation consisted on analyzing the EEG captured signals. In this case, the idea was to observe if the arousal and valence values of an image fitted with the quadrant the image was supposed to pertain. Thus, the aim of the objective evaluation was to classify the captured signals in the four quadrants. For that purpose, four machine learning algorithms were used: K-Nearest Neighbor, Decision Trees, Multilayer Perceptron and Support Vector Machines. A pre-classification step was applied in order to use the first second after the stimulus was presented to the subject. This pre-classification step was applied because the values of emotional state are more representative in the first second rather than rest [56].

3.1.6 Results

3.1.6.1 Questionnaire

The questionnaire revealed that images were selected improperly. As depicted in Figure 10 just the average of the images fitted in Q1 correlated with the selected IAPS images depicted in Figure 7. Images fitted in Q2 were lower in arousal than the expected values while images fitted in Q3 were higher in valence. Additionally, images fitted in Q4 were out of the expected quadrant.

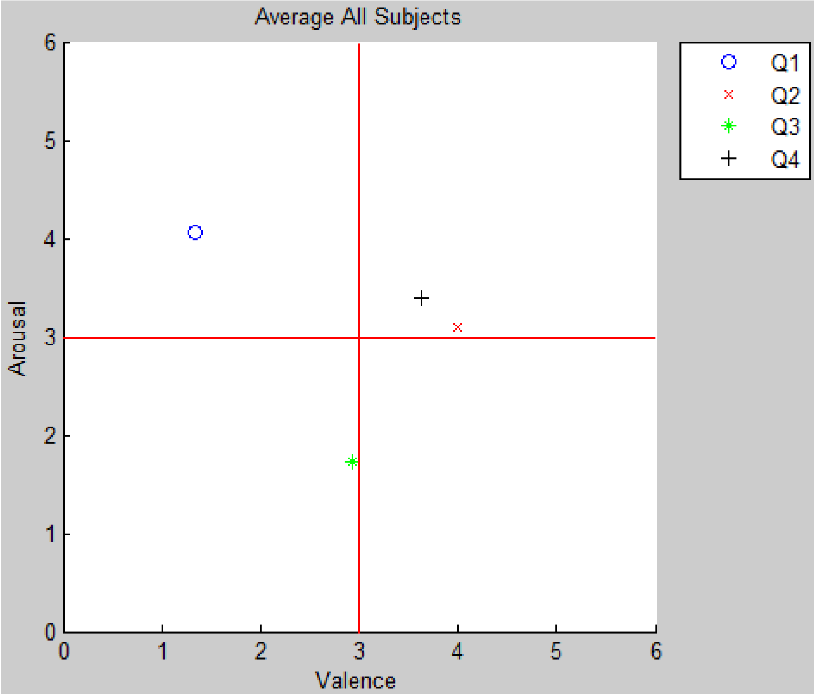


Figure 10. Average of the questionnaire answers of all subjects. The blue circle represents the average value of the three images fitted in Q1 while the red cross represents the average value of the images fitted in Q2. The green star represents the images fitted in Q3 and the black plus symbol represents the Q4 images.

3.1.6.2 EEG captured signals

Analyzing the first second after the stimuli revealed that the EEG captured signals were not correlated with the four quadrants of the arousal-valence plane. An accuracy of 35% using Multiplayer Perceptron was the best obtained result (Table 4). This is so close to the random classification (25%). The program used to classify the values was Weka [57].

Table 4. EEG captured signal values classification. Each column represents the accuracy for each subject while the last column represents the average of all subjects.

Algorithm	**Subject 1**	**Subject 2**	**Subject 3**	**Subject 4**	**Subject 5**	**Subject 6**
K-nearest neighbor	43.75%	29.17%	27.08%	31.25%	29.17%	39.58%
Decision trees	33.33%	35.42%	29.17%	39.58%	29.17%	52.08%
Multilayer Perceptron	41.67%	33.33%	22.92%	31.25%	38.00%	54.17%
Support Vector Machine	33.33%	25.00%	16.67%	25.00%	27.08%	29.17%

Algorithm	**Subject 7**	**Subject 8**	**Subject 9**	**Subject 10**	**Average**
K-nearest neighbor	39.58%	22.92%	35.42%	20.83%	31.88%
Decision trees	39.58%	25.00%	31.25%	20.83%	33.54%
Multilayer Perceptron	47.92%	25.00%	29.17%	27.08%	**35.00**%
Support Vector Machine	31.25%	22.92%	31.25%	22.92%	26.46%

3.1.7 Conclusions

Two main conclusions can be gathered from this experiment. On one hand, the improper image selection which means that other images should be selected to evoke emotions correctly. On the other hand, the no correlation between EEG captured signal values and the quadrants where the images were fitted; in other words, the EEG captured signals cannot be classified with enough accuracy.

Probably, these results are the consequence of three main reasons: (1) Image selection using the map depicted in Figure 7; (2) The images were shown during 15 seconds so many artifacts were included to the signal due to the subject eye movement or blinking. Additionally, the user had just 6 seconds to response the questionnaire and during these 6 seconds, hand movement and different brain signals were included to the captured values; and (3) Images are meant to evoke emotions but it does not mean that we feel the same our brain represents.

Due to the low accuracy obtained, another experiment was needed.

3.2 Experiment 2

The Experiment 2 was meant to avoid the mistakes made in Experiment 1. The first step was to select other images to evoke emotions since the image selection in Experiment 1 was improper. Then, in order to avoid artifacts, it was necessary to change the experiment proceeding. Software materials, Feature Extraction and Evaluation were the same than Experiment 1.

3.2.1 Image Selection

Four images were selected by hand. Image 1 and Image 4 represented High and Low arousal, respectively; while Image 2 and 3 represented Low and High valence, respectively (Figure 11).

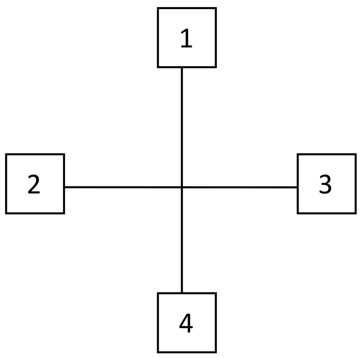


Figure 11. Image representation in the arousal-valence plane

This selection was different in terms of what the image represented. In Experiment 1, an image represented an arousal and valence value which means that one image evoked, for instance, a high and positive excitement. However, in Experiment 2, one image evoked just a high or low excitement without positive or negative connotations (Images 1 and 4) while another image evoked a positive or negative emotion without excitement connotations (Images 3 and 4). In other words, this experiment was meant to distinguish between high and low arousal as well as positive and negative valence, independently. Thus, the hypothesis is that the EEG signals corresponding to Image 1 should be higher in arousal value than the EEG signals of Image 2, as well as the EEG signals corresponding to Image 3 should be lower in valence value than the EEG signals of Image 4.

3.2.2 Experiment proceedings

3.2.2.1 Subjects

Five male subjects, members of the Music Technology Group at Universitat Pompeu Fabra, participated in the experiment. They were aged between 24 and 36.

3.2.2.2 Test

In order to avoid artifacts produced by the movement of the subject, each image lasted three seconds. In between two following images, there was a one second black image. In total, the experiment was 16 seconds long. The images were shown in the same order than is depicted in Figure 11.

As it was done in Experiment 1, the EEG signals were captured and stored in a computer in order to analyze them afterwards. Again, the captured signals were converted to discrete values using a 0.5 seconds window size and a 0.25 seconds hop-size. As a consequence, every second in time had four arousal and valence values.

3.2.2.3 Proceedings

The subject was informed to sit in front of a computer screen where the images were presented. Once the Emotiv EPOC Device was posed on the scalp of the subject, he/she was informed to not move, blink or talk during the experiment. After the experiment, the subject had to response a brief questionnaire.

3.2.2.4 Questionnaire

The plot depicted in Figure 12 was drawn four times (one per image) in a questionnaire (Appendix 2). The subject was informed to fill a point where he/she thought the image fitted better. For instance, if the subject felt that the image evoked him/her a relaxing emotion, a point had to be drawn near the word *Relax*.

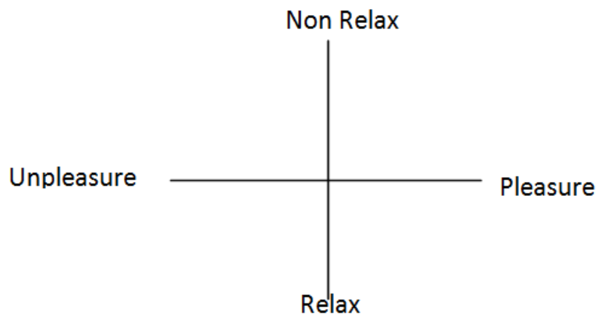


Figure 12. Plot in the questionnaire meant to fill with a point where the image fitted well according to the subject.

3.2.3 Results

The results were divided into two subsections: questionnaire results and captured EEG signals results.

3.2.3.1 Questionnaire

The results of the questionnaire revealed that the subjects agreed about the position of the images. The points which represented Image 1 and Image 4 were drawn close to *Non Relax* and *Relax* positions, respectively; while Image 2 and Image 3 were represented close to *Unpleasure* and *Pleasure*, respectively.

3.2.3.2 Captured EEG signals

The captured EEG signals were divided into two subsections as well: Average Classification and Machine Learning Classification.

Average Classification. Appendix 3 contains the EEG data per subject. The left plot depicts the average of the signal just using the first second after the stimulus was presented. The middle plot depicts the average of the two first seconds after the stimulus and the right plot depicts the average of the three seconds after the stimulus.

If one focus in the left plot of Subject 1, also depicted in Figure 13, one may observe that the EEG averaged values of Image 1 are higher in arousal than the EEG averaged values of Image 4, which agrees with the preliminary hypothesis. This happens as well with Subject 4 and Subject 5 in the first second after the stimulus.

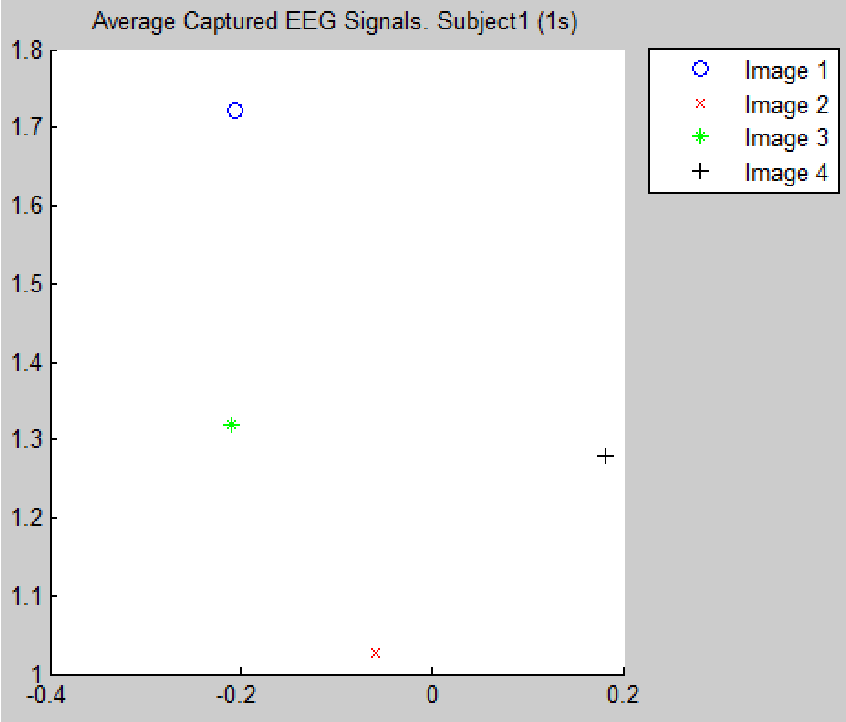


Figure 13. EEG averaged captured signals for Subject 1 in the first second after the stimulus was presented.

Instead of focusing the attention in the arousal values, one may focus on Image 2 and 3, which correspond to Low and High Valence, respectively. Figure 14 depicts the EEG averaged signals for Subject 2 in the three following seconds after the stimulus was presented. As one can observe, Image 2 and 3 are in the left side and in the right side, one respect to the other, which means that the EEG averaged values of Image 2 were lower in valence than the EEG averaged values of Image 3. The same happens with Subject 4 and Subject 5.

Average classification demonstrates that the division between High and Low arousal as well as Low and High valence is feasible. In order to make this affirmation stronger, machine learning classification algorithms were used.

Machine Learning Classification. Instead of classifying the data using four classes, one per image; the EEG captured data was divided into Arousal (High and Low) and Valence (Low and High). Subject 3 was discarded because his/her brain signals did not follow the preliminary hypothesis.

Table 5 shows the arousal classification applying four machine learning techniques. The best accuracy was achieved by using K-nearest neighbor classification algorithm with K=1 (75%) over the values obtained during the first second after the stimulus. Table 6 shows the results using the three seconds after the stimulus. The results revealed that the accuracy when classifying arousal values is better when using the first second after the stimulus rather than using the three seconds after.

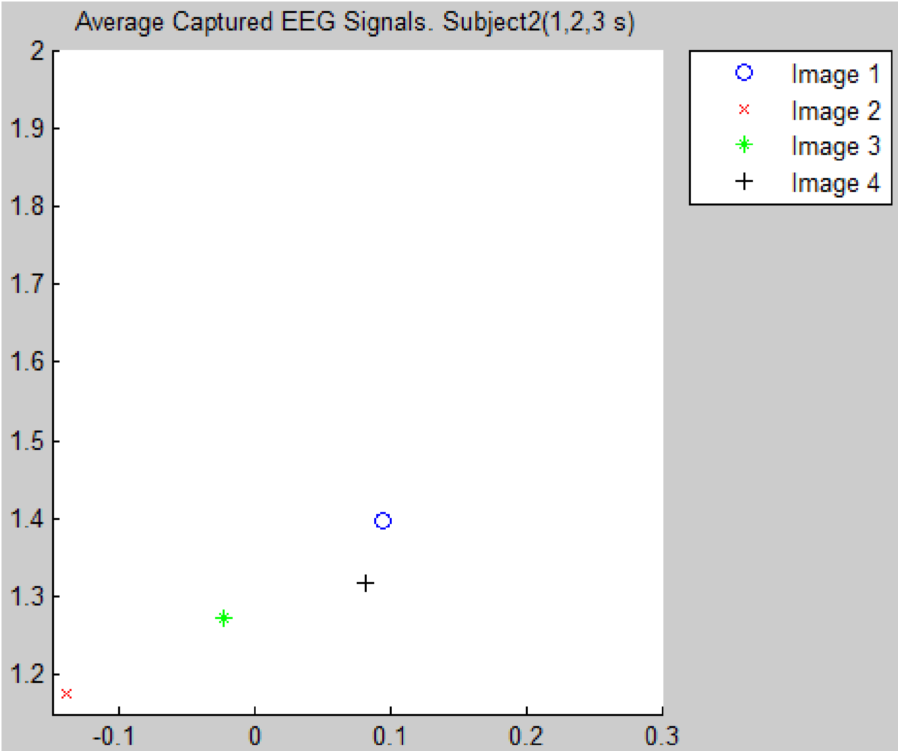


Figure 14. EEG averaged captured signals for Subject 2 in the three following seconds after the stimulus was presented.

Table 5. Arousal classification using the first second after the stimuli. Each classification algorithm used 8 fold cross validation to train the model.

Algorithm	Subject 1	Subject 2	Subject 4	Subject 5	Average
K-nearest neighbor	37.50%	75%	87.50%	100%	**75.00%**
Decision trees	75.00%	0%	62.50%	87.50%	56.25%
Multilayer Perceptron	50.00%	37.50%	75.00%	100%	65.63%
Support Vector Machine	12.50%	12.50%	25.00%	87.50%	34.38%

Table 6. Arousal classification using the three seconds after the stimuli. Each classification algorithm used 22 fold cross validation to train the model.

Algorithm	Subject 1	Subject 2	Subject 4	Subject 5	Average
K-nearest neighbor	45.45%	36.36%	77.27%	59.09%	**54.55%**
Decision trees	72.73%	0%	81.82%	40.91%	48.86%
Multilayer Perceptron	63.64%	4.55%	72.73%	59%	50.00%
Support Vector Machine	45.45%	4.55%	81.82%	0%	32.95%

Table 7 shows the valence classification applying the same machine learning techniques. The best accuracy was achieved by using K-nearest neighbor classification algorithm with K=1 (71.88%) over the three seconds after the stimulus. Table 8 shows the results using the first second after the stimulus. In this case, the valence classification had more accuracy when using the three seconds after the stimuli rather than using just the first second.

Table 7. Valence classification using the three seconds after the stimuli. Each classification algorithm used 22 fold cross validation to train the model.

Algorithm	Subject 1	Subject 2	Subject 4	Subject 5	Average
K-nearest neighbor	62.50%	87.50%	100%	37.50%	**71.88%**
Decision trees	0%	87.50%	87.50%	62.50%	59.38%
Multilayer Perceptron	75.00%	75.00%	100%	50%	75.00%
Support Vector Machine	25.00%	25.00%	87.50%	12.50%	37.50%

Table 8. Valence classification using the first second after the stimuli. Each classification algorithm used 8 fold cross validation to train the model.

Algorithm	Subject 1	Subject 2	Subject 4	Subject 5	Average
K-nearest neighbor	40.91%	54.55%	45.45%	72.73%	**53.41%**
Decision trees	31.82%	0%	22.73%	45.45%	25.00%
Multilayer Perceptron	59.09%	40.91%	59.09%	82%	60.23%
Support Vector Machine	45.45%	13.64%	22.73%	0%	20.45%

3.2.4 Conclusions

The main conclusion extracted from the questionnaire results was that the images were selected properly. The first mistake done in Experiment 1, image selection, was solved in Experiment 2. On the other hand, in Experiment 2 the subjects introduced fewer artifacts than in Experiment 1 due to the selected proceedings. Instead of answering the questionnaire during the experiment, the subjects answered the questions after it. It avoided artifacts regarding hand and head movements. Additionally, 3 seconds per image was enough time because after this amount of seconds the concentration of the subject diminished.

The average classification results demonstrated that the difficulty trying to classify images fitted in the four quadrants of the arousal-valence plane decreases when classifying arousal and valence independently. In Experiment 1, arousal values introduced noise when trying to classify valence values and conversely. In Experiment 2, this mistake was also avoided.

Machine learning techniques gave a proper accuracy classifying arousal and valence values. 75% (arousal case) and 71.88% (valence case) are higher than the random accuracy (50%). This machine learning classification made stronger the average classification conclusion.

Moreover, one observe that gathering arousal values is better in the first second after the stimulus while gathering valence values is better when using the three seconds after the stimulus. The extracted conclusion from these results is that the arousal can be immediately extracted after the stimulus while the valence needs more time. Probably, this is because arousal means instant state of mind and saying if an emotion is positive or negative needs few seconds more.

3.2.5 Comparison with other experiments

The experiments described in [6], [7] and [10] were the basis of Experiments 1 and 2 of this master thesis. The aim of this section is to compare the experiments of the state-of-the-art and Experiment 2 to see the main differences and similarities between them.

Table 9 depicts the main items of each experiment. Oude [6] and Choppin [7] centered their research in finding the best formulas to define valence and arousal values while Vamvakousis' study [10] and Experiment 2 were focused on finding the best way to classify arousal and valence using fixed formulas. As a consequence, [6] and [7] used multiple feature extraction combinations.

Emotiv EPOC neuroheadset was used in [10] and Experiment 2 while [6] and [7] used more sophisticated devices. Regarding the montage and brainwaves, electrodes F3 and F4 and frequencies alpha and beta were used in all the experiments.

The test proceedings showed a similar way to present the stimuli which consisted of presenting the stimulus during few seconds and introducing a black image in between two followed stimuli to have a rest. The amount of seconds per stimulus varied depending on the experiment. [6] and [7] used audio, visual and audiovisual stimuli while [10] used audio stimuli and in Experiment 2 just visual stimuli were used.

In the classification process, except in [7], all the experiments used machine learning techniques to classify the data and the best results in terms of accuracy were obtained by Oude achieving more than a 90% in both arousal and valence classifications.

In conclusion, the main differences between experiments where the goal to achieve (two of them found the best formulas and the others tried to find the best classification) and the classification process because all of them used different techniques (machine learning or manual). Additionally, the main similarities were that the experiments used IAPS, IADS or both; the test proceeding; the used brainwaves to detect emotions and the use of prefrontal and frontal electrodes.

Table 9. Comparison between the experiments done in the state-of-the-art using IAPS and IADS stimuli and Experiment 2.

	Oude [6]	**Choppin[7]**	**Vamvakousis [10]**	**Experiment 2**
Subjects	5	20	6	5
Neuroheadset	BraInquiry PET 2.0	Digital Bio-Amplifier 5200	Emotiv EPOC	Emotiv EPOC
Electrodes	F3,F4 and Fpz	Fp1, Fp2, F3, F4, F7, F8, T3, T4, P3, P4, O1and O2	AF3, AF4, F3, F4	AF3, AF4, F3, F4
Frequencies	Alpha and beta	Alpha and beta	Alpha and beta	Alpha and beta
Feature Extraction	alpha, beta, alpha and beta, beta / alpha, alpha power, beta power, alpha and beta power, beta / alpha power, and beta power / alpha power	Combination of alpha and beta using: Log PSD, Log power asymmetry, Coherence, Peak frequency, Cross-bicoherence	Arousal: beta/alpha ratio Valence: AF4/BF4 – AF3/BF3	Arousal: beta/alpha ratio Valence: AF4/BF4 – AF3/BF3
Materials	EEGLab for MatLab	MatLab	OpenVibe	OpenVibe, MatLab, Weka
Stimuli	IAPS and IADS	IAPS and IADS	IADS	IAPS
Stimuli selection	Audio, visual and audiovisual stimuli (36 in total)	32 pictures, 32 sounds and 30 combinations	12 extreme sound stimuli	4 image stimuli
Test proceedings	Five seconds of stimuli exposure will be followed by ten seconds of cooling down	Five seconds of warning message, 6 or 10 seconds of stimulus exposure and 15 seconds of rating	Five seconds of stimulus and 10 second silent rest is inserted between stimuli	Three seconds image stimulus followed by one second black image
Classification	PCA and FDA	Rating gives the separation of classes	LDA and SVM	K-nearest neighbor, J48, MultiLayer Perceptron and SVM
Results	Arousal 97.4%. Valence: 94.9%	Arousal: High beta power and coherence in parietal lobe and low alpha activity Valence: high frontal alpha coherence and high parietal beta power (Positive)	Arousal: 77.82%, Valence: 80.11	Arousal: 75%, Valence: 71.88%

3.3 Emotion-based BCMI approach

In order to answer the research questions, a device which uses emotions as inputs was implemented. The main idea was to connect the captured EEG signals and the expressive music performance using a self-paced BCMI (Figure 15) and to demonstrate the hypothesis that high arousal captured signals are higher in value than low arousal captured signals.

3.3.1 Previous considerations

As demonstrated in Experiment 2, valence needs at least three seconds to be well classified in Low and High while arousal can be immediately calculated. As a consequence, the control of the application was done by taking into account just the arousal values because the Emotion-based BCMI was meant to be a real-time application.

Another consideration should be the use of images to calculate the boundaries of arousal values. In Experiment 1 they did not work due to its improper selection while in Experiment 2, the selected images obtained just a 75% of accuracy when classifying High and Low arousal values. Nevertheless, Sergio Giraldo and Rafael Ramirez used another approach for controlling a BCMI which consisted in instructing the subject to be excited or relaxed by thinking of events, images, situations or whatever that makes the user feel these two opposite states of mind [58]. Instead of using images, this experiment used this last approach because it fitted better with the purpose of the Emotion-based BCMI: to separate High and Low arousal values.

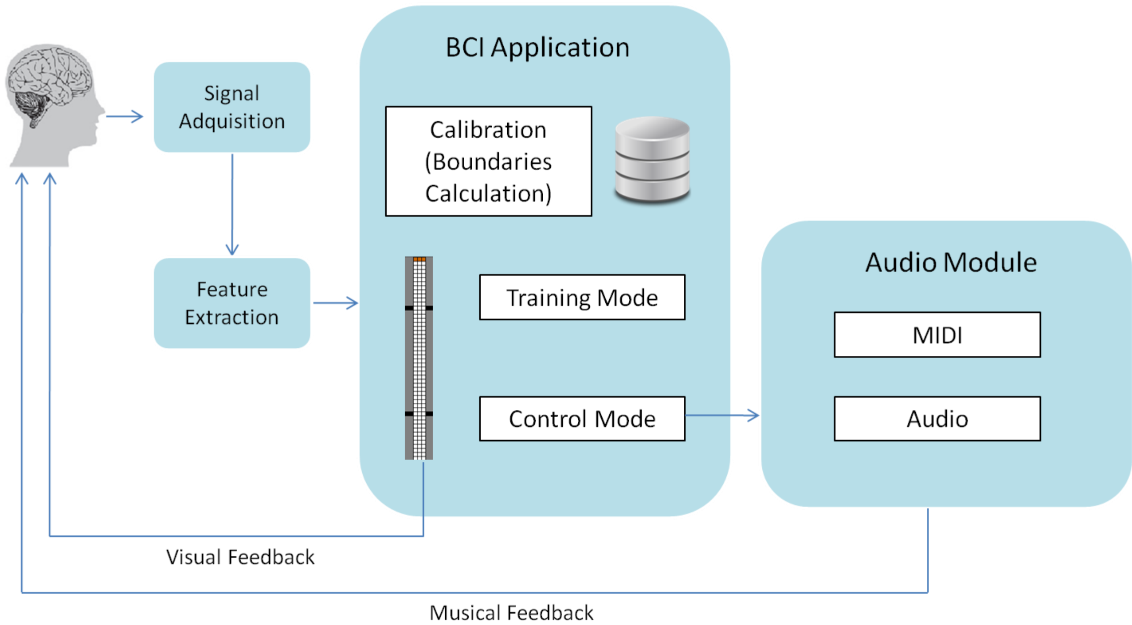


Figure 15. Parts of the Emotion-based BCMI.

3.3.2 Materials

Emotiv EPOC device and OpenViBE were used again for signal acquisition and feature extraction, respectively. Processing[2] is a software development environment mostly used for prototyping and building interactive programs. It can be used to create user interfaces and connect the brain signal values received from OpenViBE software with PureData [51], a software environment which allows changing the expressive performance of an excerpt that is played in real-time.

3.3.3 Emotion-based BCMI design

Hereinafter the Emotion-based BCMI design is explained.

3.3.3.1 Signal Acquisition and Feature Extraction

The main feature to take into account was the arousal state of mind. OpenViBE received the signal from the Emotiv EPOC device and processed it using a 4 seconds window-size and a 1 second hop-size. It means that every second one arousal value was obtained and sent to the BCI application.

Firstly, a spatial filtering selected the electrodes to be used (AF3, AF4, F3 and F4). Then, a band pass filter was applied to the signal in order to acquire α and β waves. Afterwards, a power and logarithmic formulas converted the values in positive ones. Finally, the arousal values were obtained by applying a β/α ratio (Eq2).

3.3.3.2 BCI Application

This part of the BCMI was build using Processing and connected to OpenVibe (Feature Extraction) using VRPN to OSC converter

- **Calibration (Boundaries Calculation).** The boundaries calculation was meant to determinate which maximum and minimum arousal values a subject can achieve. A visual feedback consisting on a grid with an orange bar was drawn on a screen as depicted in Figure 16. In order to fit the position of the orange bar an initial mapping was necessary. The calibration boundaries ranged from 0 to 2 (approximately the minimum and maximum values gathered from the Experiments 1 and 2). Additionally, two arrows were depicted to instruct the subject to be excited or relaxed depending on the direction the arrows pointed at. Up arrow meant excite state of mind while down arrow meant relax state of mind.

- **Training Mode.** Training Mode depicted a grid with an orange bar and two limits in order to offer visual feedback to the subject (Figure 17). The orange bar reached the top of the grid value when the arousal value was equal or higher than the determined maximum boundary in the Calibration process. Conversely, when the arousal value was equal or lower than the determined minimum boundary, the

[2] http://www.processing.org/

orange bar reached the bottom of the grid. In other words, the mapping between EEG signals and visual feedback was scaled using the boundaries calculated in the Calibration process.

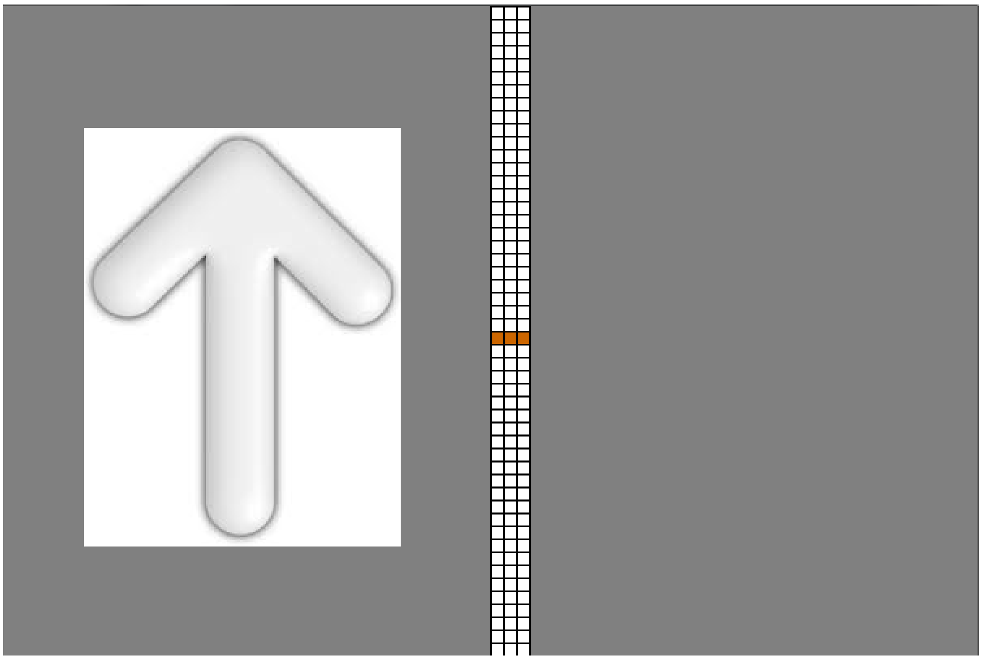


Figure 16. Visual feedback shown as a grid with an orange bar. The arrow is depicted in order to guide the subject in the calibration process. When up arrow appears, the subject tries to be as excited as possible. Conversely, when down arrow appears, the subject tries to be as relax as possible.

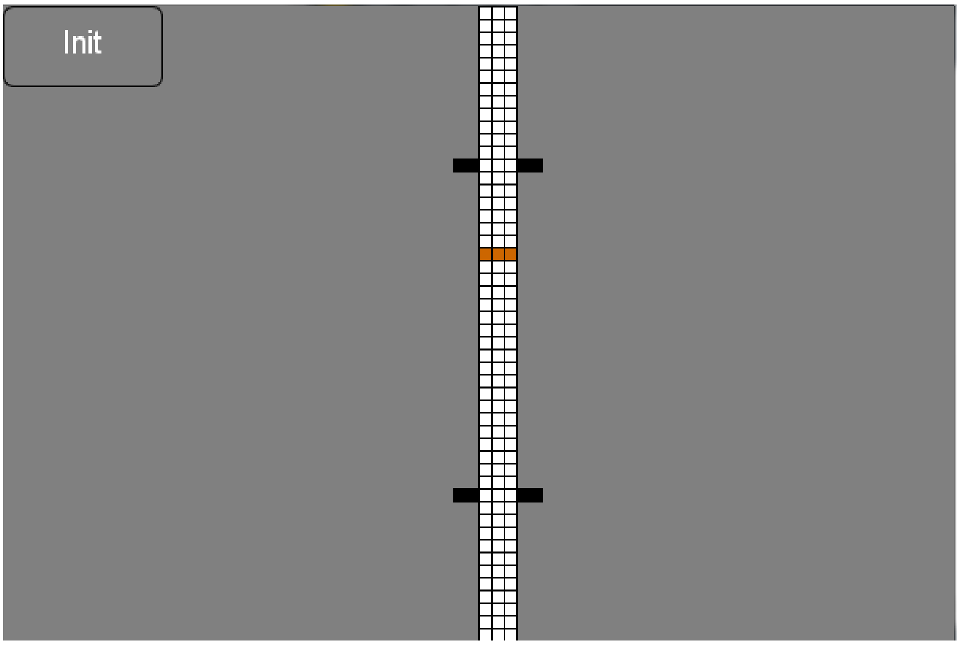


Figure 17. Visual feedback depicted in Training Mode.

- **Control Mode.** Each arousal value was again scaled depending whether the excerpt to control was MIDI or polyphonic audio. In the former case the arousal values were mapped between 0 and 400 while, in the latter, these values were mapped in the range from 1 to 127. Afterwards, these values were sent to the Audio Module.

3.3.3.3 Audio Module

- **MIDI.** Audio module changes the expressive performance of a MIDI excerpt using pDM (Programs to control music performance through emotions). This program used a grid were a point was moved depending on the values received from the Control Mode. In Figure 18 the grid is depicted with the point in position (200,200). Instead of changing just the arousal values, the program also changed the valence values in order to increase the expressive performance. As a consequence, the point did not move in vertical direction, it moved tracing a diagonal going from Happy to Sad and conversely.

  The mapped values were sent from the BCI Application to the Audio Module via OSC protocol.

- **Polyphonic audio.** An online time stretching algorithm [52] and volume control program were used to control a polyphonic audio excerpt. As a consequence, two musical features were used: speed and loudness. When the arousal value was high, the music played louder and faster while, if the arousal values were lower, the expressive performance changed to slow and soft.

  The time stretching algorithm and volume control were implemented in a VST-plugin [59] and connected to the BCI application using PureData and MIDI control protocol.

3.3.4 Experiment Proceedings

Eleven subjects participated in the experiment, five females and six males. They were aged between 23 and 48. Each subject was instructed to sit in front of a screen where the BCI application was depicted (Figure 19). Firstly, the application had to be calibrated and then the subject trained in order to move an orange bar up and down by achieving an exciting or relaxing state of mind.

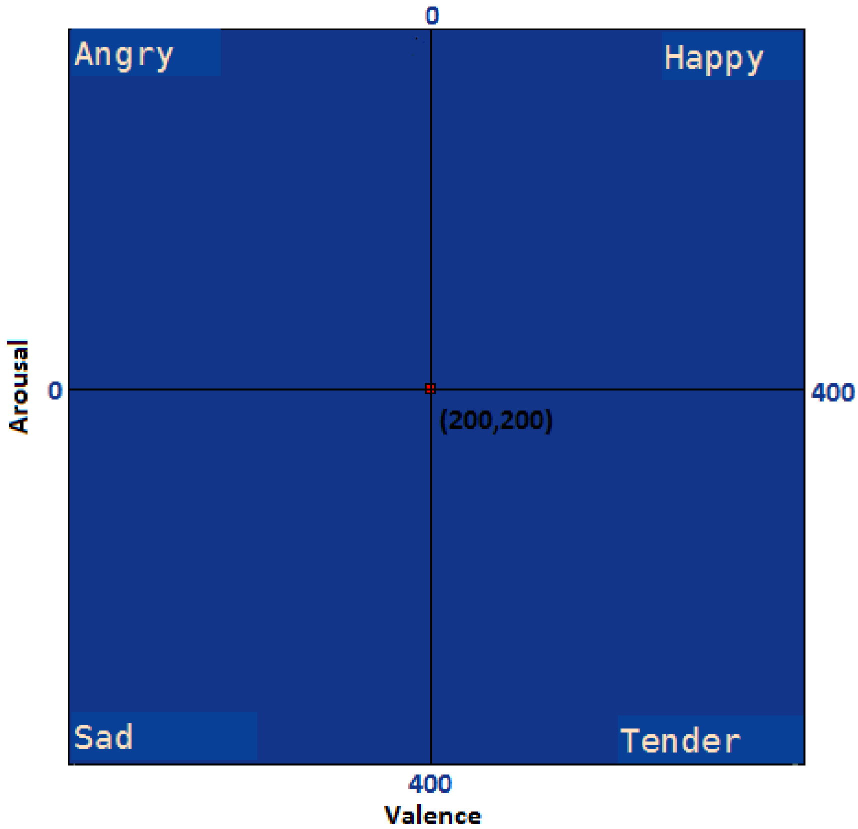


Figure 18. 2D arousal-valence space used in pDM to change the expressive music performance using the KTH rule system.

Secondly, the application was calibrated again in order to start a *visual feedback scenario*. At this scenario the user could see the grid with the orange bar and was instructed to be excited or relaxed. Then, the subject was meant to control the music in the *visual and musical feedback scenario*. At this scenario, a MIDI excerpt was played and the subject was instructed to be excited to make the music play louder and faster or to be relaxed in order to make the music play softer and slower. Additionally, a polyphonic audio excerpt was played and the same states of mind were expected.

After the aforementioned scenario case, the experiment was repeated with the eyes closed. Two more scenarios were recorded: *no feedback scenario* and *musical feedback scenario*. The application was calibrated again because the brain signals are different when having the eyes closed or opened. In the *no feedback scenario* the subject did not receive any visual or musical feedback; however, during *the musical feedback scenario* there was no visual feedback so the subject just listened to the music (first MIDI and then polyphonic audio excerpt).

Once the experiment finalized, each subject answered a questionnaire in order to evaluate the application subjectively (Appendix 4).

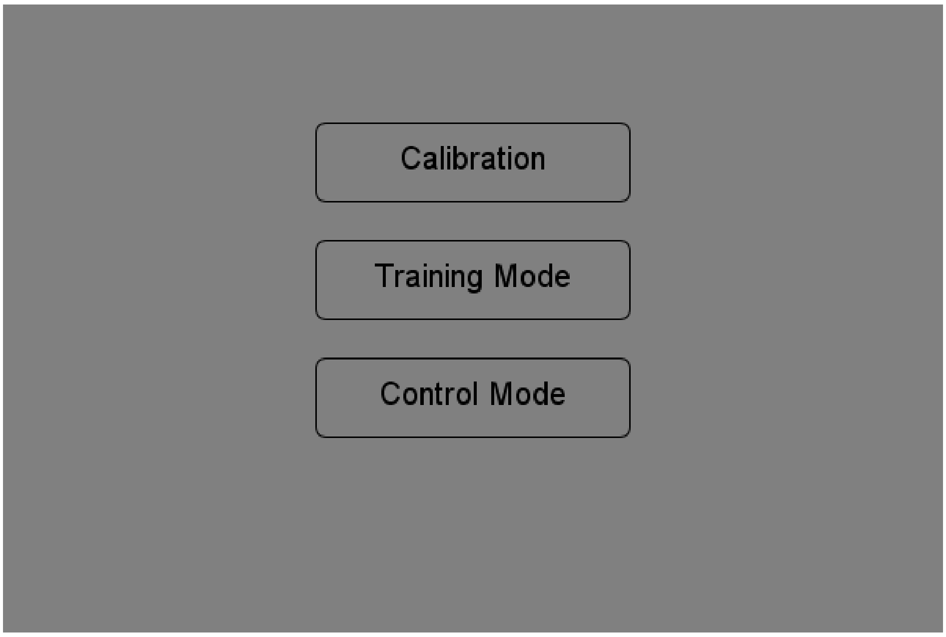


Figure 19. BCI Application first screen. The button 'Calibration' started the boundaries calculation process. 'Training Mode' button showed the grid with the orange bar and 'Control Mode' button allowed to control the musical excerpts.

3.3.5 Evaluation

Six scenarios for each subject were evaluated: (1) visual feedback, (2) visual and musical feedback using MIDI excerpt, (3) visual and musical feedback using polyphonic audio excerpt, (4) no feedback, (5) musical feedback using MIDI excerpt and (6) musical feedback using polyphonic audio excerpt.

For each scenario 15-seconds high and low arousal periods were recorded. Figure 20 depicts the high (blue) and low (red) arousal values captured from Subject 5 corresponding to musical feedback scenario using polyphonic audio excerpt and, in order to give a measurement, it also depicts the mean of each period. The same procedure was done for each scenario (Figure 21). Finally, the average of the means of all scenarios was calculated in order to obtain one value for high arousal and one for low arousal (Figure 22). This approach was called subject-oriented evaluation because it measured two arousal values (high and low) for each subject. Appendix 5 contains all the graphs regarding subject-oriented evaluation.

In addition, another approach called scenario-oriented was meant to evaluate one scenario taking into account all the subjects. Firstly, the means of high and low arousal periods for each subject were calculated (Figure 23), obtaining one value for high arousal and another one for low arousal when calculating the average of all means (Figure 24). Appendix 6 contains all the graphs regarding scenario-oriented evaluation.

Additionally, the questionnaire offered a qualitative and subjective evaluation in order to determine the level of control the subjects felt when using the application. Every subject answered three questions regarding three scenarios: visual feedback scenario, musical and visual feedback scenario and musical feedback scenario. Each question ranged from 1 to 5, being 1 the lower feeling of control and 5 the highest.

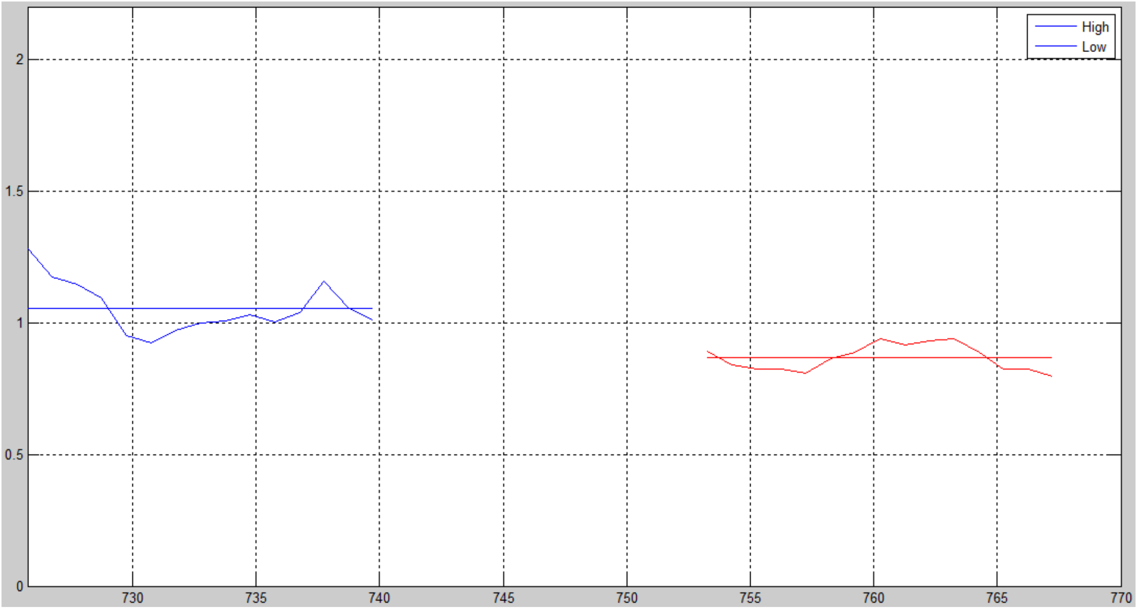


Figure 20. High arousal period (in blue) and low arousal period (in red) captured from Subject 5's brain signals during musical feedback scenario using polyphonic audio excerpt. Each period lasts 15 seconds. Blue flat line and red flat line represents the mean of high and low arousal periods, respectively.

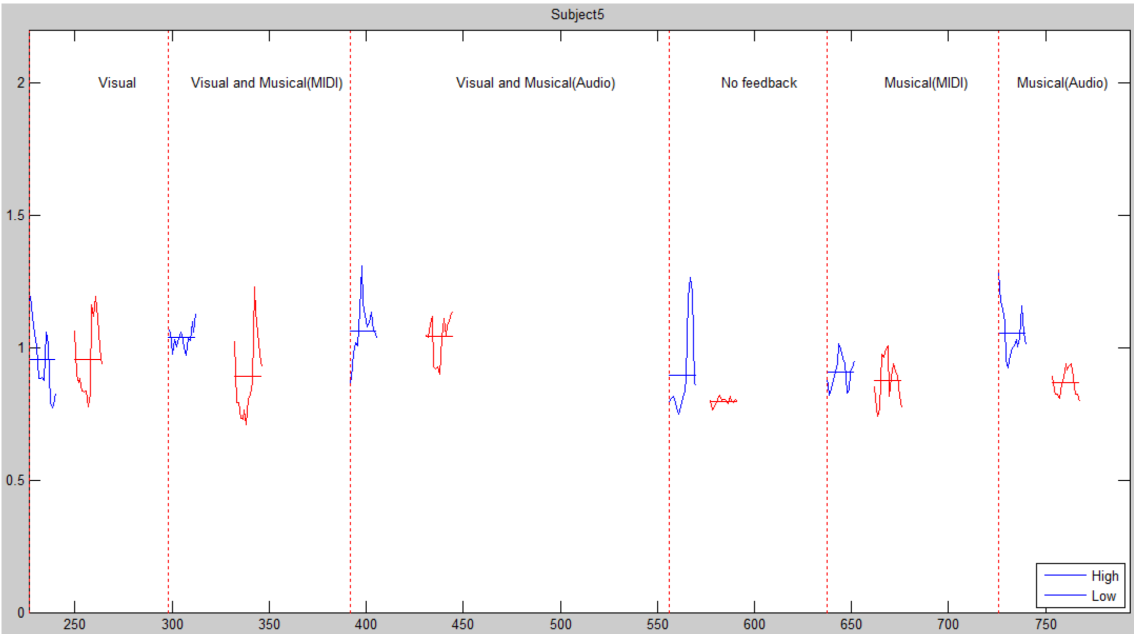


Figure 21. High (blue) and low (read) arousal periods for each scenario. Flat lines are the mean values for each period. This plot corresponds to Subject 5's brain signals.

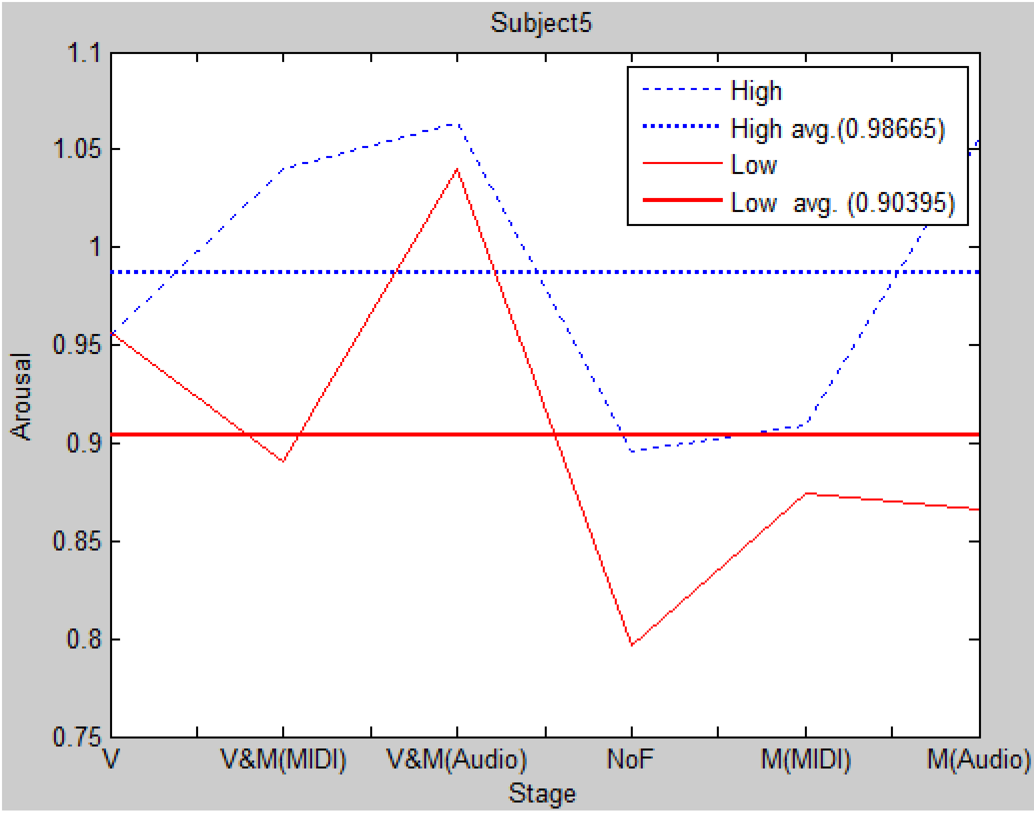


Figure 22. High (blue) and low (red) arousal means for each scenario and the total high and low averages concerning Subject 5. The scenarios are Visual (V), Visual and Musical (V&M), No feedback (NoF) and Musical (M). 'MIDI' means a MIDI excerpt was played while 'Audio' means a polyphonic audio excerpt was played.

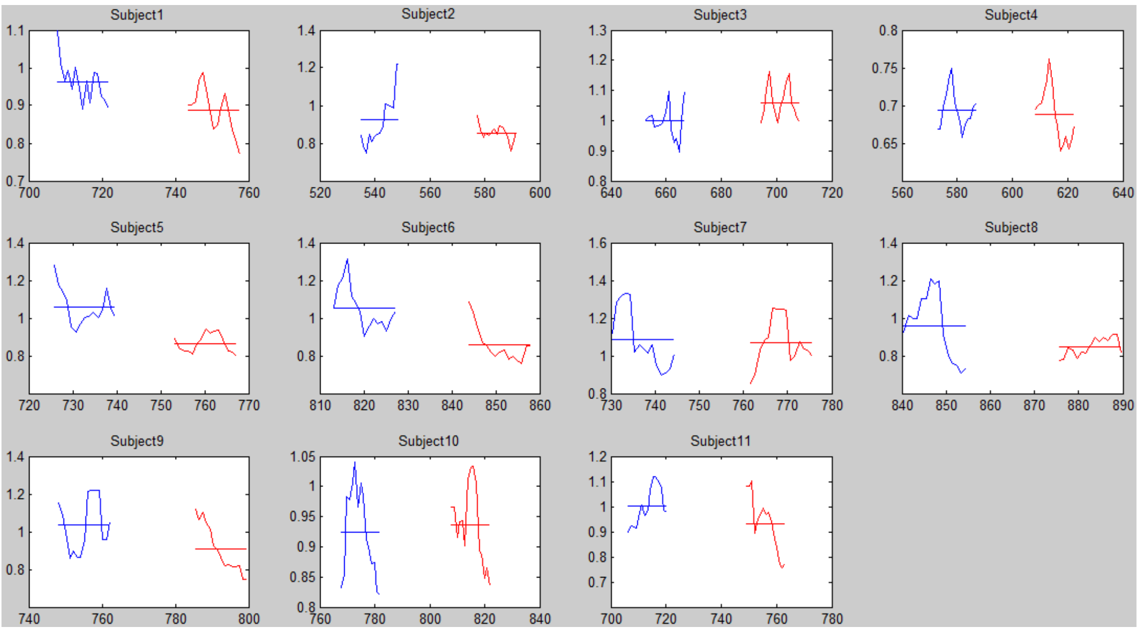


Figure 23. High (blue) and low (red) arousal periods and their mean values (flat lines) for all subjects during musical scenario using polyphonic audio excerpt.

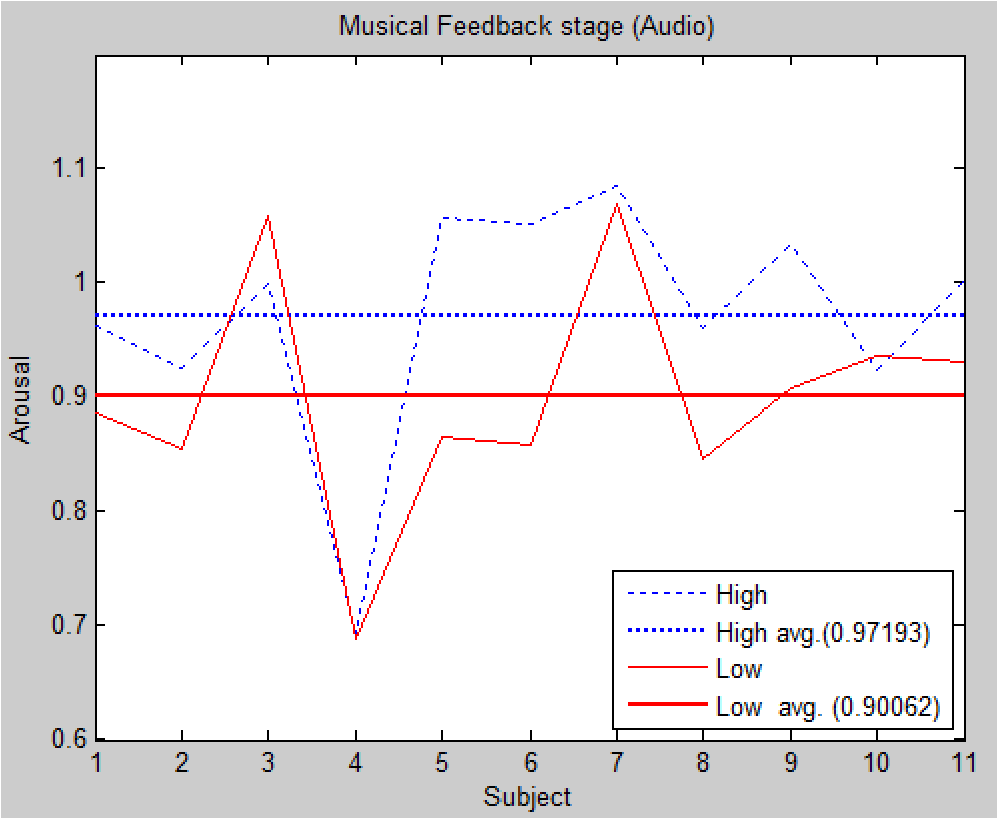


Figure 24. High (blue) and low (red) mean values per subject and their average values (flat lines).

3.3.6 Results

Table 10 depicts the high and low arousal mean values for each subject in all scenarios (user-oriented evaluation) while the mean values of high and low arousal for each scenario using the average of all subjects (scenario-oriented evaluation) are given in Table 11.

Table 10. High and Low arousal averages for each subject calculated from each individual scenario mean value.

Subject	High Arousal avg.	Low Arousal avg.
1	1.1455	1.1184
2	0.99585	0.97276
3	1.1285	1.1671
4	0.81784	0.8088
5	0.98665	0.90395
6	0.95589	0.93873
7	1.0876	0.99811
8	1.0956	1.1318
9	0.97725	0.89229
10	0.99207	0.98692
11	1.1533	1.0492

Table 11. High and low arousal averages for each scenario calculated from each subject mean value.

Scenario	High Arousal avg.	Low Arousal avg.
Visual feedback	1.0859	1.1205
Visual and Musical feedback (MIDI)	1.1293	1.0758
Visual and Musical feedback (Audio)	1.1051	1.0998
No feedback	0.93233	0.88958
Musical feedback (MIDI)	0.95878	0.89637
Musical feedback (Audio)	0.97193	0.90062

Two subjects (3 and 8) and one scenario (Visual feedback) did not accomplish the preliminary hypothesis: when the subject was supposed to be excited the brain signals should be higher in arousal than when the subject was relaxed.

Regarding questionnaire results, subjects felt more control using musical feedback rather than visual feedback. The feeling of control during visual and musical feedback scenario remained in between the aforementioned scenarios.

3.3.7 Conclusions

The user-oriented evaluation revealed that nine of the eleven subjects were able to do the task properly. It means that the average of the mean values of each task was higher in arousal value when the subject was supposed to be excited, and lower when he/she was supposed to be relaxed.

Regarding scenario-oriented evaluation, just visual feedback did not fulfill the hypothesis. Probably, this happened because the scenario was early in the experiment and the subject had not trained enough to control his/her brain signals.

The questionnaire results revealed that the feeling of control was stronger when giving the subject musical feedback.

Thus, the answers to the research questions are:

- ***Is it possible to control expression in music through emotions using BCIs?***

  When using the arousal values (excited and relaxed emotions) the expressive music performance can be controlled and changed at will. In conclusion, brain signals can be used to control a BCMI.

- ***Can the user easily control an Emotion-based BCMI?***

  When having just musical feedback (eyes closed) the user felt more control than having the eyes opened and receiving visual feedback. Thus, the user easily controlled an Emotion-based BCMI receiving musical feedback.

3.3.8 Repeating the experiment

To repeat the experiment using the Emotiv EPOC neuroheadset, one should install the Emotiv EPOC Research Edition Development Kit (http://emotiv.com/store/sdk/bci/research-edition-sdk/) in which the TestBench is included. To reproduce the experiment one should follow the next steps:

- Download and install OpenVibe from http://openvibe.inria.fr/
- Download and install Processing from https://processing.org/download/
- Download and install PureData from http://puredata.info/downloads
- Install a VST-plugin host program that uses MIDI control protocol. In that experiment AudioMulch (http://www.audiomulch.com/) was used.
- Open the TestBench of the Emotiv EPOC Research Edition and be sure that the brain signals arrive to the computer
- Open the OpenVibe Acquisition Server and select "Emotiv EPOC" in Driver input. After that press Connect and Play buttons.
- Open the OpenVibe Designer and open this file http://www.weebly.com/uploads/1/8/4/7/18476064/bcmiexperiment.xml
- Download this file http://www.weebly.com/uploads/1/8/4/7/18476064/vrpn2osc.rar, open the .rar file and execute the .exe. This allows the communication between OpenVibe and Processing.
- Open Processing and open the file http://www.weebly.com/uploads/1/8/4/7/18476064/bcmi.pde. To make sure that Processing works using OCS protocol one should install the library "oscP5" (http://www.sojamo.de/libraries/oscP5/)
- Open PureData and open the file http://www.weebly.com/uploads/1/8/4/7/18476064/osc_udpreceive.pd which controls a MIDI file that should be selected in the same Pure Data program. Make sure that you select an output driver to reproduce MIDI.
- Open another Pure Data and open the file http://www.weebly.com/uploads/1/8/4/7/18476064/audiocontrol.pd. Make sure that you select an output MIDI driver to send commands to the VST-plugin host program (e.g. LoopBe1 - http://www.nerds.de/en/loopbe1.html)
- Open the VST-plugin host program and open this VST-plugin: http://www.weebly.com/uploads/1/8/4/7/18476064/vstitempocontrol.dll. After that, select a .VAW file to control. Make sure that you selected the input MIDI driver to receive orders from Pure Data.
- Start the execution of the current scenario in OpenVibe Designer
- Start the execution of the Processing program (See how to control this program in Emotion-based BCMI design)
- If one selects to control MIDI, press Play in the Pure Data program which controls a MIDI file. If one selects to control Audio, press Play in the VST-plugin host that control a .WAV file.

In order to analyze the extracted data, these two programs written in MatlLab were used: http://www.weebly.com/uploads/1/8/4/7/18476064/matlab.rar. `BCMI.m` was used for user-oriented evaluation while `BCMIPart.m` was used for scenario-oriented evaluation.

4. Contributions and Future Work

4.1 Contributions

Three contributions can be extracted from this thesis. First, the arousal and valence calculations used in [10] are also valid by using image stimuli. This confirms that these two formulas can be used as well using audio or picture stimulus. Second, arousal values can be calculated immediately but valence needs more time to be calculated with enough accuracy. One reason may be that valence means to evaluate something positively or negatively and it requires more time than determining if something has a high or low impact (arousal). And third, the Arousal-driven BCMI is a contribution in itself because it tries to fill the gap between BCI and Emotion and Expressive Music Performance using emotions. In addition, this new application can be used to do more experiments and still contributing to the research community.

4.2 Future Work

In future Emotion-based applications the whole arousal-valence plane can be covered if valence is included. As a consequence, more emotions can be used to change the expression of music. On the other hand, instead of using prefrontal cortex electrodes one can think of using electrodes posed on the motor cortex. This will provide a new set of inputs for BCMIs to control, create or compose music.

References

[1] B. Blankertz, M. Krauledat, G. Dornhege, J. Williamson, R. Murray-Smith, and K.-R. Müller, “A Note on Brain Actuated Spelling with the Berlin Brain-Computer Interface,” *Universal Access in HumanComputer Interaction Ambient Interaction*, vol. 4555, pp. 759–768, 2007.

[2] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. D. R. Millán, “A brain-actuated wheelchair: asynchronous and non-invasive Brain-computer interfaces for continuous control of robots.,” *Clinical Neurophysiology*, vol. 119, no. 9, pp. 2159–2169, 2008.

[3] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, “Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic.,” *Computational intelligence and neuroscience*, vol. 2007, p. 8, Jan. 2007.

[4] K. Oum, H. Ayaz, P. A. Shewokis, and P. Diefenbach, “MindTactics: A Brain Computer Interface gaming platform,” in *Games Innovations Conference ICEGIC 2010 International IEEE Consumer Electronics Societys*, 2010, pp. 1–5.

[5] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. Ramsey, I. Sturm, G. Curio, and K.-R. Müller, “The Berlin Brain–Computer Interface: Non-Medical Uses of BCI Technology,” *Frontiers in neuroscience*, vol. 4, no. December, p. 17, Jan. 2010.

[6] D. Oude Bos, “EEG-based Emotion Recognition The Influence of Visual and Auditory Stimuli,” *Emotion*, vol. 57, no. 7, pp. 1798–1806, 2006.

[7] A. Choppin, “EEG-Based Human Interface for Disabled Individuals: Emotion Expression with Neural Networks,” Yokohama, Japan, 2000.

[8] K. Takahashi, “Remarks on emotion recognition from multi-modal bio-potential signals,” in *Industrial Technology, 2004. IEEE ICIT ’04. 2004 IEEE International Conference on*, 2004, vol. 3, pp. 1138–1143 Vol. 3.

[9] Y.-P. Lin, C.-H. W. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, “EEG-based emotion recognition in music listening.,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[10] R. Ramirez and Z. Vamvakousis, “Detecting Emotion from EEG Signals Using the Emotive Epoc Device,” in in *Brain Informatics SE - 17*, vol. 7670, F. Zanzotto, S. Tsumoto, N. Taatgen, and Y. Yao, Eds. Springer Berlin Heidelberg, 2012, pp. 175–184.

[11] A. Lucier, “Music for Solo Performer.” 1965.

[12] E. R. Miranda, “Plymouth brain-computer music interfacing project: from EEG audio mixers to composition informed by cognitive neuroscience,” *International Journal of Arts and Technology*, vol. 3, no. 2/3, pp. 154–176, 2010.

[13] E. R. Miranda, W. L. Magee, J. J. Wilson, J. Eaton, and R. Palaniappan, “Brain-Computer Music Interfacing (BCMI): From Basic Research to the Real World of Special Needs,” *Music and Medicine*, vol. 3, no. 3, pp. 134–140, Mar. 2011.

[14] E. R. Miranda and A. Brouse, “Toward Direct Brain-Computer Musical Interfaces,” in *in 2005 Conference on New Interfaces for Musical Expression (NIME05)*, 2005.

[15] P. Juslin and P. Laukka, “Communication of emotions in vocal expression and music performance: different channels, same code?,” *Psychological Bulletin*, vol. 129, no. 5, pp. 770–814, Sep. 2003.

[16] P. Juslin and P. Laukka, “Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening,” *Journal of New Music Research*, vol. 33, no. 3, pp. 217–238, Sep. 2004.

[17] A. Friberg, R. Bresin, and J. Sundberg, “Overview of the KTH rule system for musical performance,” *Advances in Cognitive Psychology*, vol. 2, no. 2, pp. 145–161, 2006.

[18] R. Bresin and A. Friberg, “Emotion rendering in music: range and characteristic values of seven musical variables.,” *Cortex; a journal devoted to the study of the nervous system and behavior*, vol. 47, no. 9, pp. 1068–1081, Oct. 2011.

[19] M. Fabiani, “Interactive computer-aided expressive music performance. Analysis, control, modification, and synthesis methods,” KTH School of Computer Science and Communication, Stockholm, Sweden, 2011.

[20] A. Friberg, “pDM : An Expressive Sequencer with Real-Time Control of the KTH Music- Performance Rules,” *Computer Music Journal*, vol. 30, no. 1, pp. 37–48, 2006.

[21] N. Moriguchi, E. Wada, and M. Miura, “Emotion control system for MIDI excerpts: MOR2ART,” in *Proceedings of the 3rd International Conference on Fun and Games*, 2010, pp. 58–65.

[22] J. J. Vidal, “Toward direct brain-computer communication.,” *Annual review of biophysics and bioengineering*, vol. 2, pp. 157–180, Jan. 1973.

[23] K. G. Oweiss, *Statistical signal processing for neuroscience and neurotechnology*. Elsevier, 2010.

[24] E. A. Larsen and A. I. Wang, “Classification of EEG Signals in a Brain- Computer Interface System,” 2011.

[25] A. Subhani, L. Xia, and A. Malik, "EEG Signals to Measure Mental Stress," in *2nd International Conference on Behavioral, Cognitive and Psychological Sciences - BCPS 2011*, 2011, p. 10.

[26] M. a Lebedev and M. a L. Nicolelis, "Brain-machine interfaces: past, present and future.," *Trends in neurosciences*, vol. 29, no. 9, pp. 536–546, Sep. 2006.

[27] B.-K. Min, M. J. Marzelli, and S.-S. Yoo, "Neuroimaging-based approaches in the brain-computer interface.," *Trends in biotechnology*, vol. 28, no. 11, pp. 552–560, Nov. 2010.

[28] Trans Cranial Technologies Ltd., "10 / 20 System Positioning Manual." Trans Cranial Technologies ldt., Wanchai, Hong Kong, p. 20, 2012.

[29] N. Birbaumer, "Breaking the silence: brain-computer interfaces (BCI) for communication and motor control.," *Psychophysiology*, vol. 43, no. 6, pp. 517–532, Nov. 2006.

[30] B. Blankertz, M. Krauledat, G. Dornhege, J. Williamson, R. Murray-Smith, and K.-R. Müller, "A note on brain actuated spelling with the Berlin brain-computer interface," *Lecture Notes in Computer Science*, vol. 4557, pp. 759–768, 2007.

[31] D. Stokes and M. S. Lappin, "Neurofeedback and biofeedback with 37 migraineurs: a clinical outcome study," *Behavioral and brain functions BBF*, vol. 6, no. 1, p. 9, 2010.

[32] J. A. Pineda, D. S. Silverman, A. Vankov, and J. Hestenes, "Learning to control brain rhythms: making a brain-computer interface possible," *Neural Systems and …*, vol. 11, no. 2, pp. 181–184, 2003.

[33] H.-J. Hwang, K. Kwon, and C.-H. Im, "Neurofeedback-based motor imagery training for brain-computer interface (BCI).," *Journal of neuroscience methods*, vol. 179, no. 1, pp. 150–156, Apr. 2009.

[34] T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49, Aug. 2010.

[35] J. A. Russell, "A circumplex model of affect.," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[36] M. Bradley and P. Lang, "International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings," Gainesville, FL, 1999.

[37] P. Lang, M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," Gainesville, FL, 2008.

[38] R. Teitelbaum, "In Tune: Some Early Experiments in Biofeedback Music," in in *Biofeedback and the Arts, Results of Early Experiments*, D. Rosenboom, Ed. Vancouver: Aesthetic Research Centre of Canada, 1975, pp. 35–56.

[39] D. Rosenboom, *Extended musical interface with the human nervous system : assessment and prospectus*. Berkeley, CA.: Internationhal Society for the Arts, Sciences and Technology, 1990.

[40] A. Overton, "Sitting.Breathing.Beating.[Not]Thinking for solo performer," in *Conference Proceedings of Consciousness Reframed 6: QI and COMPLEXITY*, 2004.

[41] T. Hinterberger, "ORCHESTRAL SONIFICATION OF BRAIN SIGNALS AND ITS APPLICATION TO BRAIN-COMPUTER INTERFACES AND PERFORMING ARTS Thilo Hinterberger Division of Social Sciences , University of Northampton , UK Institute of Medical Psychology and Behavioral Neurobiology , Uni," in *Proceedings of the 2nd International Workshop on Interactive Sonification*, 2007.

[42] T. Hinterberger, "The Sensorium: A Multimodal Neurofeedback Environment," *Advances in Human-Computer Interaction*, vol. 2011, pp. 1–10, 2011.

[43] D. Wu, C. Li, Y. Yin, C. Zhou, and D. Yao, "Music composition from the brain signal: representing the mental state by music.," *Computational intelligence and neuroscience*, vol. 2010, p. 6, Jan. 2010.

[44] B. Hamadicharef, M. Xu, and S. Aditya, "Brain-Computer Interface (BCI) Based Musical Composition," in *2010 International Conference on Cyberworlds*, 2010, pp. 282–286.

[45] Y. C. (Denise) Chew and E. Caspary, "MusEEGk," in *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, 2011, pp. 1417–1422.

[46] E. R. Miranda, "Brain-Computer Interface for Generative Music," in *Proceedings of International Conference Series on Disability Virtual Reality and Associated Technologies ICDVRAT 2006*, 2006, pp. 1–8.

[47] Z. Vamvakousis and R. Ramirez, "A Brain-Gaze Controlled Musical Interface," in *Berlin BCI Workshop 2012 - Advances in Neurotechnology*, 2012.

[48] W. Goebl, S. Dixon, G. De Poli, A. Friberg, R. Bresin, and G. Widmer, "Sense in expressive music performance: Data acquisition, computational studies, and models," in in *Sound to Sense - Sense to Sound: A state of the art in Sound and Music Computing*, D. Polotti, Pietro; Rocchesso, Ed. Berlin, Germany: Logos Verlag, 2008, pp. 195–242.

[49] R. Ramirez and A. Hazan, “A Tool for Generating and Explaining Expressive Music Performances of Monophonic Jazz Melodies,” *International Journal on Artificial Intelligence Tools*, vol. 15, pp. 673–691, 2006.

[50] S. I. Giraldo and R. Ramirez, “Modeling Embellishment , Duration and Energy Expressive Transformations in Jazz Guitar,” 2012.

[51] M. Puckette, “Pure Data: another integrated computer music environment,” in *in Proceedings, International Computer Music Conference*, 1996, pp. 37–41.

[52] J. Bonada, J. Janer, and S. Jordà, “Groovator - An Implementation of Real-Time Rhythm Transformations,” in *Audio Engineering Society Convention 121*, 2006.

[53] M. Fabiani, “A method for the modification of acoustic instrument tone dynamics,” in *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-09)*, 2009.

[54] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer, “OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments,” *Presence Teleoperators and Virtual Environments*, vol. 19, no. 1, pp. 35–53, 2010.

[55] T. H. Aspiras and V. K. Asari, “Log power representation of EEG spectral bands for the recognition of emotional states of mind,” *2011 8th International Conference on Information, Communications & Signal Processing*, pp. 1–5, Dec. 2011.

[56] M. M. Bradley, S. Hamby, A. Löw, and P. J. Lang, “Brain potentials in perception: picture complexity and emotional arousal.,” *Psychophysiology*, vol. 44, no. 3, pp. 364–73, May 2007.

[57] E. Frank, M. Hall, and G. Holmes, “Weka-a machine learning workbench for data mining,” *Data Mining and …*, pp. 1–11, 2010.

[58] S. Giraldo and R. Ramirez, “Brain-Activity-Driven Real-Time Music Emotive Control,” in *… 3rd International Conference on Music & Emotion (…*, 2013, no. June.

[59] O. Mayor, J. Bonada, and J. Janer, “Audio Transformation Technologies Applied to Video Games,” in *41st AES Conference: Audio for Games*, 2011.

Appendix 1

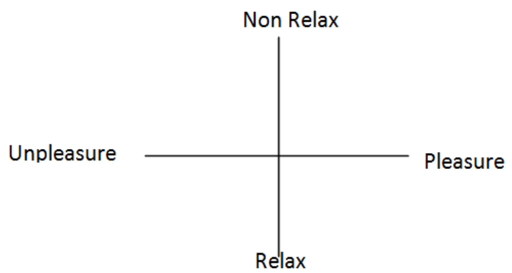
EMOTION CLASSIFICATION EXPERIMENT

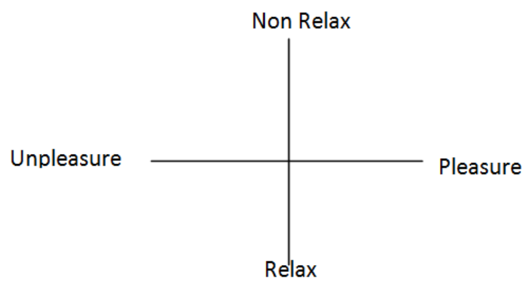
Age: Gender: Male/Female

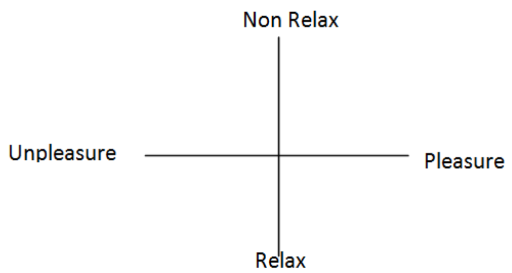
	Arousal/Excitement					**Valence/Pleasure**				
	MIN				MAX	−				+
1	1	2	3	4	5	○	○	○	○	○
2	1	2	3	4	5	○	○	○	○	○
3	1	2	3	4	5	○	○	○	○	○
4	1	2	3	4	5	○	○	○	○	○
5	1	2	3	4	5	○	○	○	○	○
6	1	2	3	4	5	○	○	○	○	○
7	1	2	3	4	5	○	○	○	○	○
8	1	2	3	4	5	○	○	○	○	○
9	1	2	3	4	5	○	○	○	○	○
10	1	2	3	4	5	○	○	○	○	○
11	1	2	3	4	5	○	○	○	○	○
12	1	2	3	4	5	○	○	○	○	○

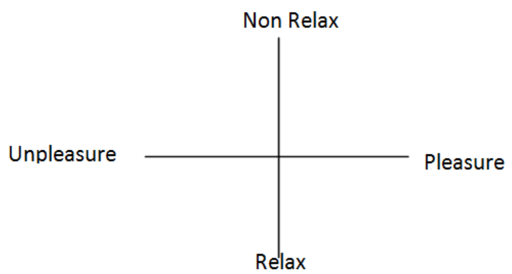
Appendix 2

Age: Male/Female



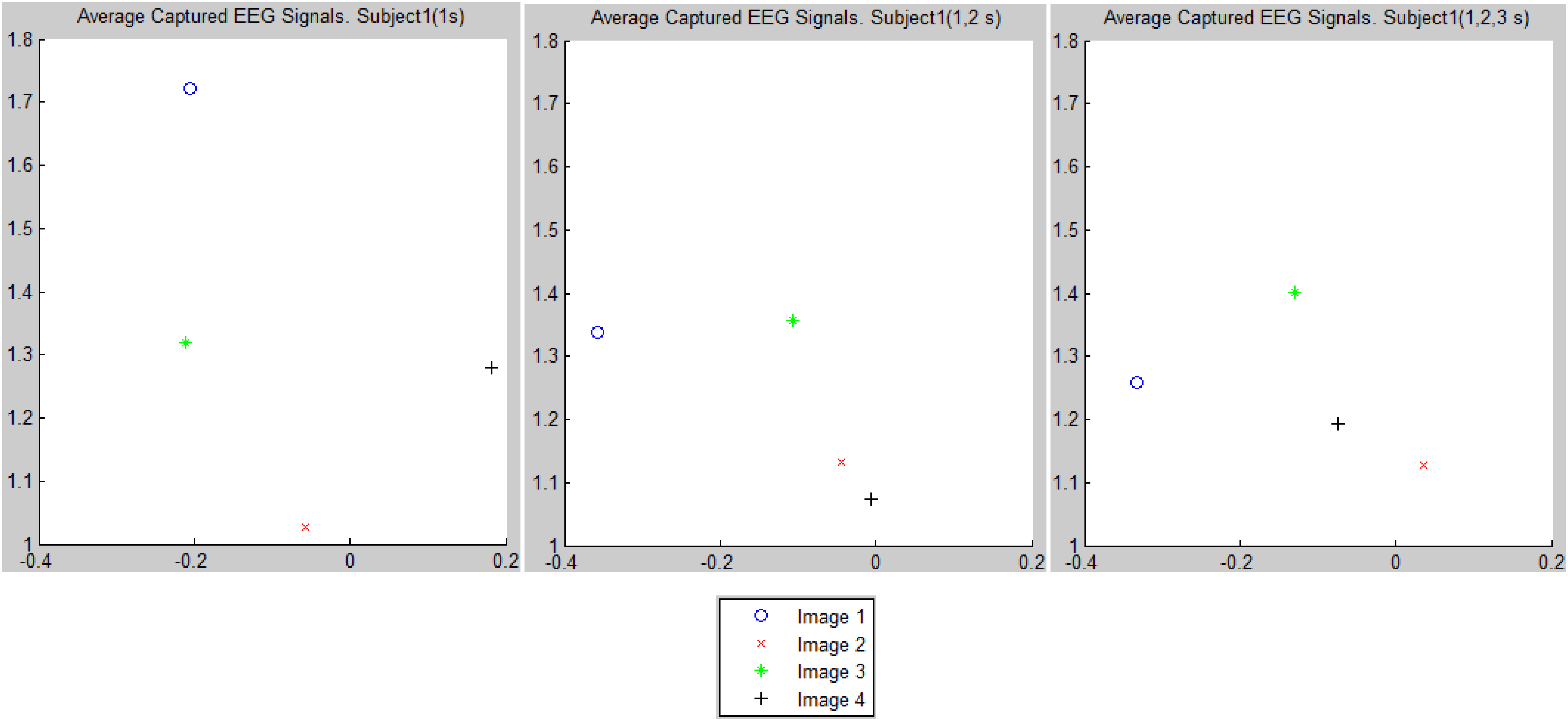




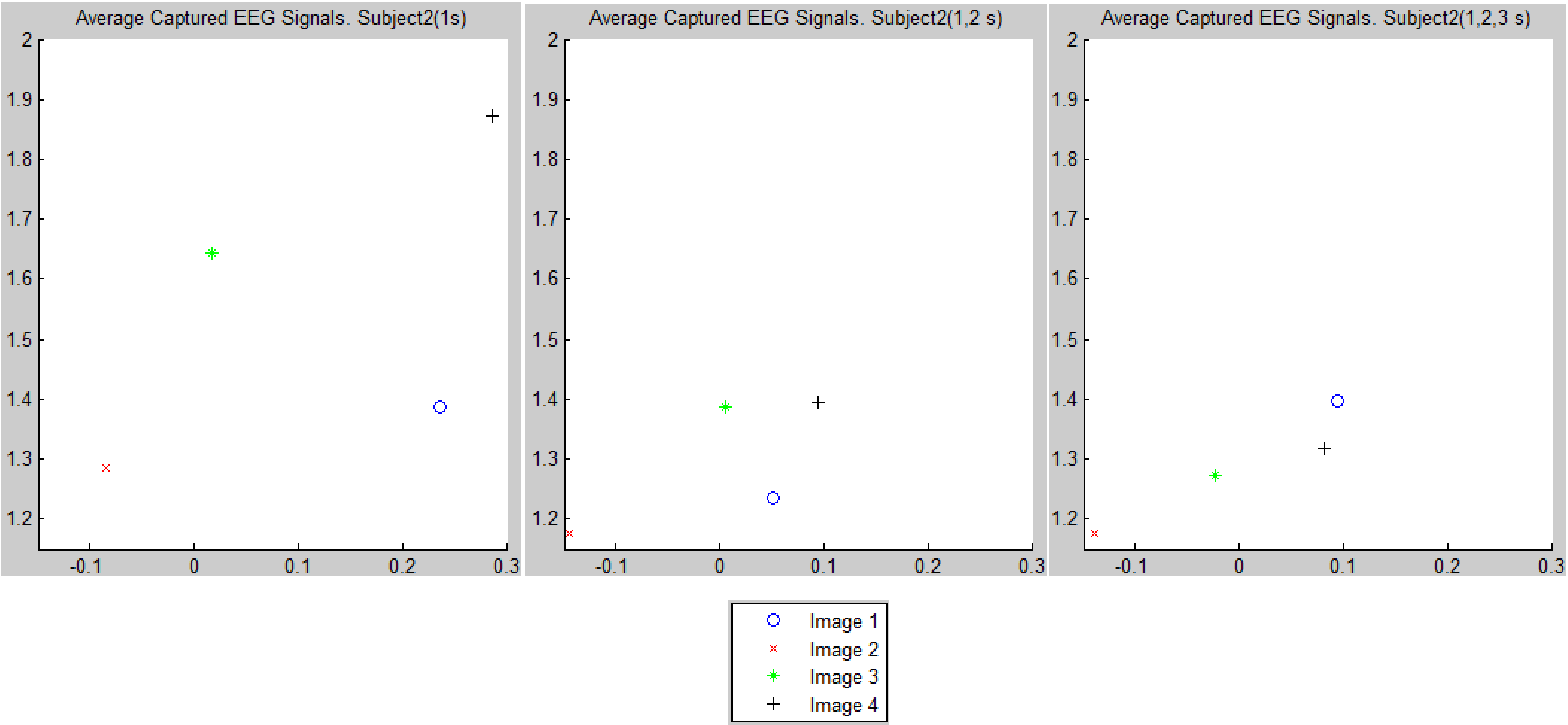


Appendix 3

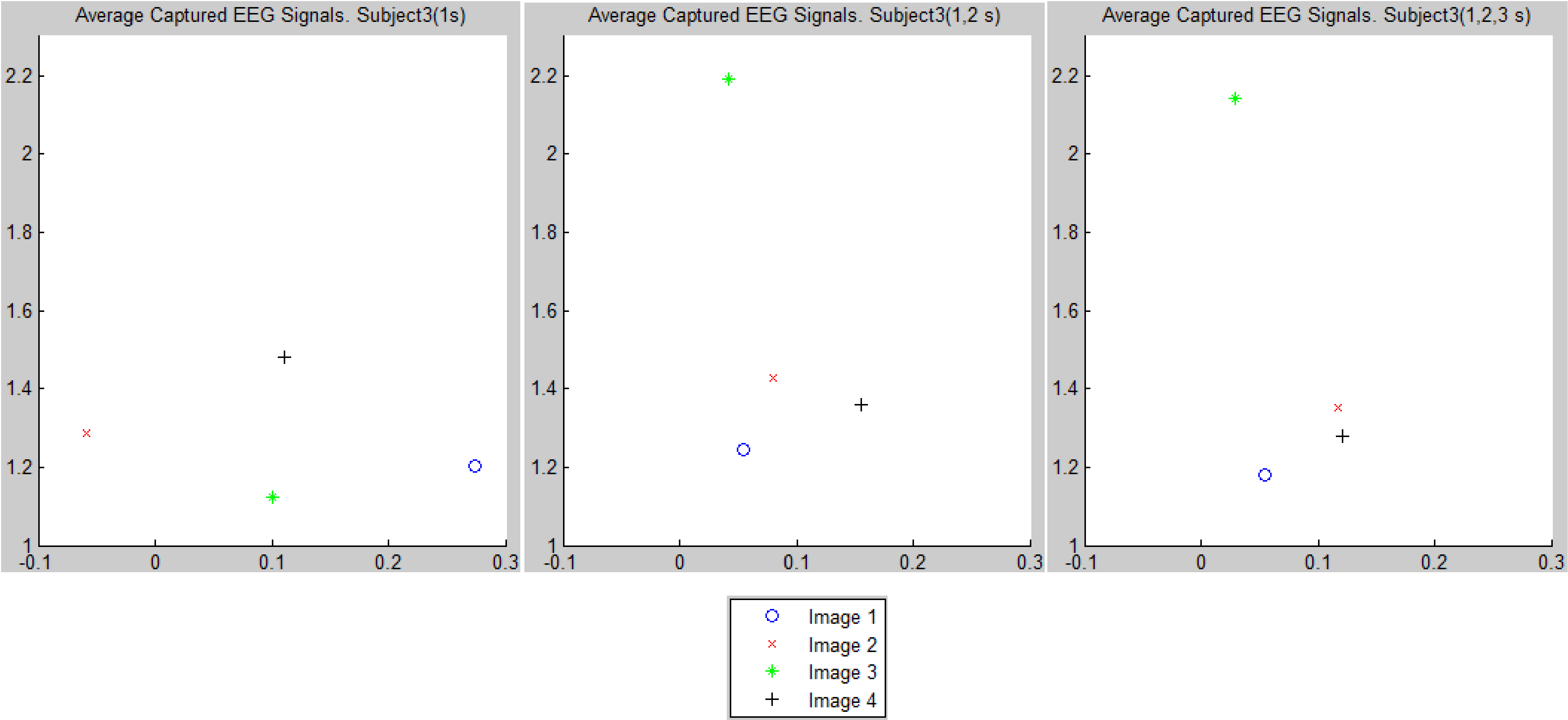
Subject 1



Subject 2

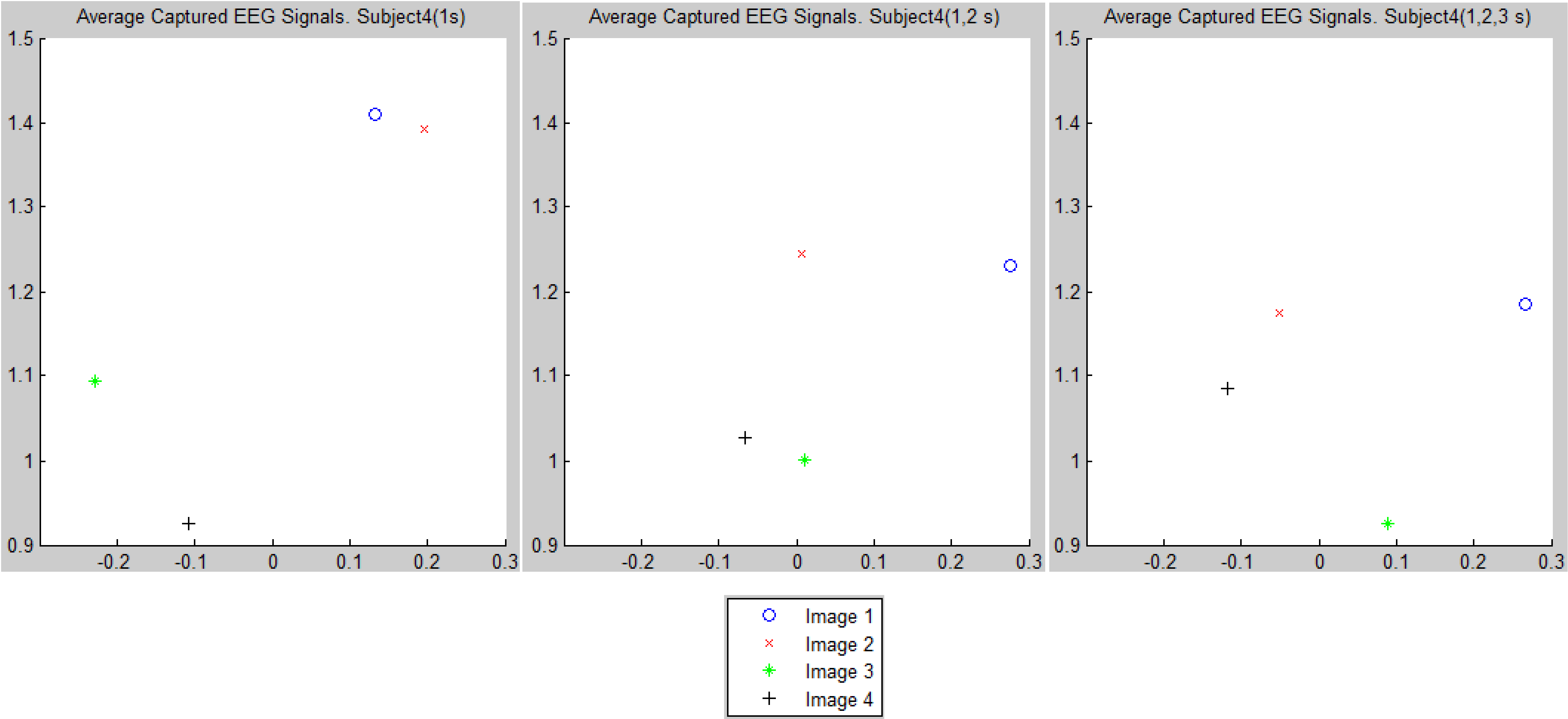

Average Captured EEG Signals. Subject2(1s)
Average Captured EEG Signals. Subject2(1,2 s)
Average Captured EEG Signals. Subject2(1,2,3 s)
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
-0.1
0
0.1
0.2
0.3
Image 1
Image 2
Image 3
Image 4

Subject 3



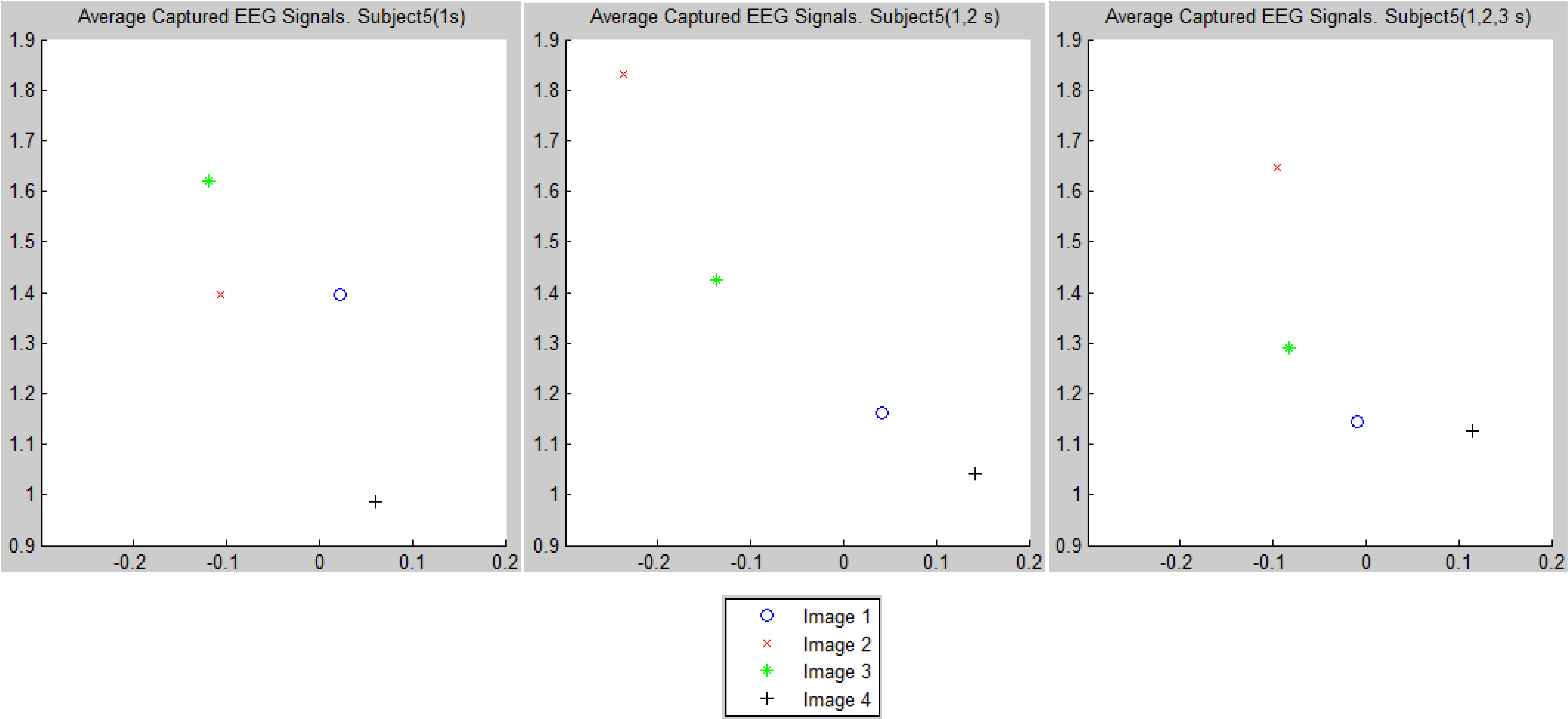
Average Captured EEG Signals. Subject3(1s)
Average Captured EEG Signals. Subject3(1,2 s)
Average Captured EEG Signals. Subject3(1,2,3 s)
Image 1
Image 2
Image 3
Image 4

Subject 4



Average Captured EEG Signals. Subject4(1s)
Average Captured EEG Signals. Subject4(1,2 s)
Average Captured EEG Signals. Subject4(1,2,3 s)
1.5
1.4
1.3
1.2
1.1
1
0.9
-0.2
-0.1
0
0.1
0.2
0.3
Image 1
Image 2
Image 3
Image 4

Subject 5



Average Captured EEG Signals. Subject5(1s)
Average Captured EEG Signals. Subject5(1,2 s)
Average Captured EEG Signals. Subject5(1,2,3 s)
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
-0.2
-0.1
0
0.1
0.2
Image 1
Image 2
Image 3
Image 4

Appendix 4

EMOTION-BASED BCMI EXPERIMENT

Subject number ___

Age:

Genre: Male/Female

QUESTIONS

1. Did you feel control using visual feedback?

MIN				MAX
1	2	3	4	5

2. Did you feel control using visual and musical feedback?

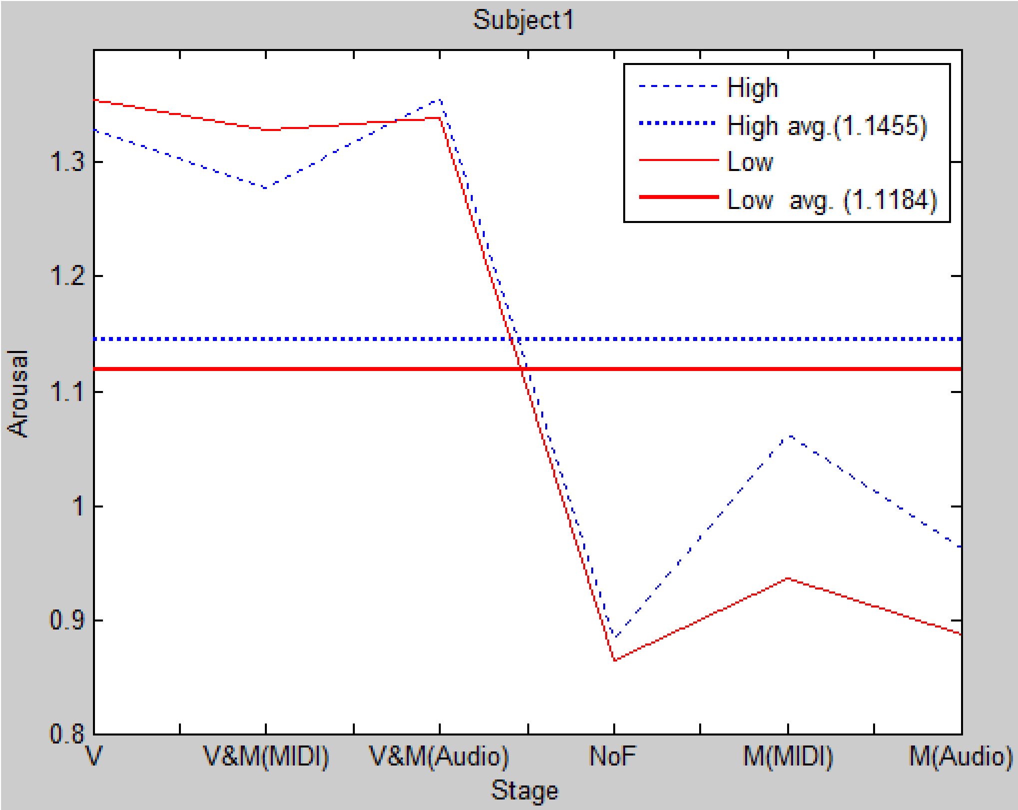
MIN				MAX
1	2	3	4	5

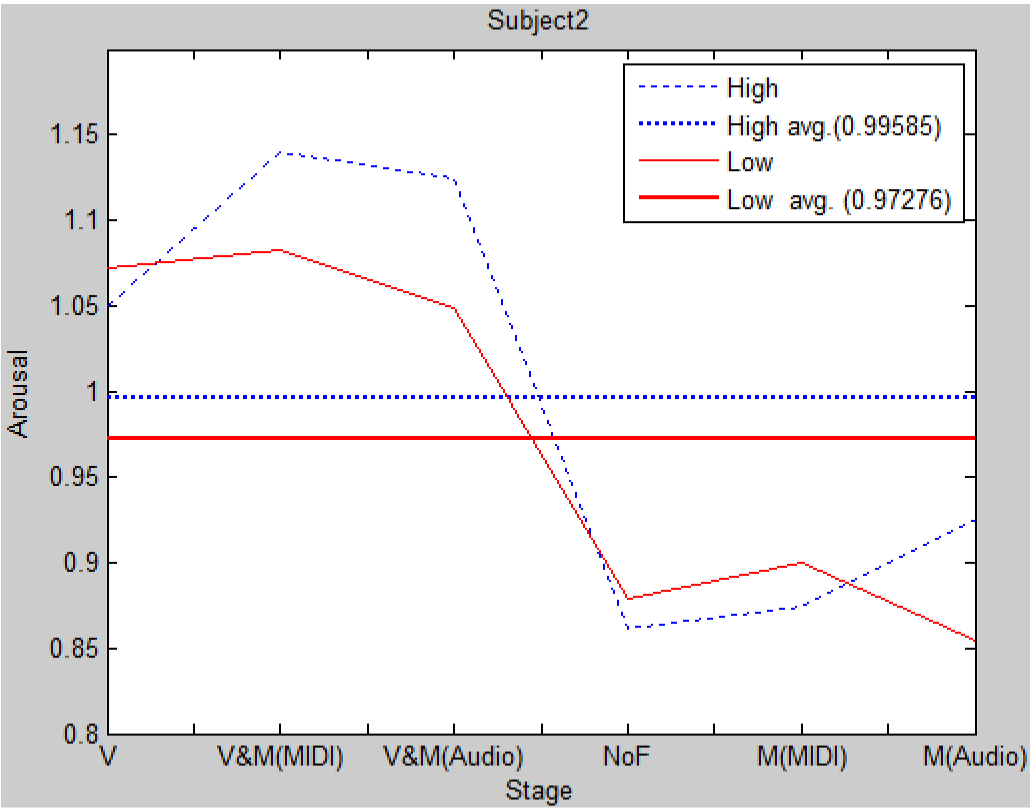
3. Did you feel control using musical feedback?

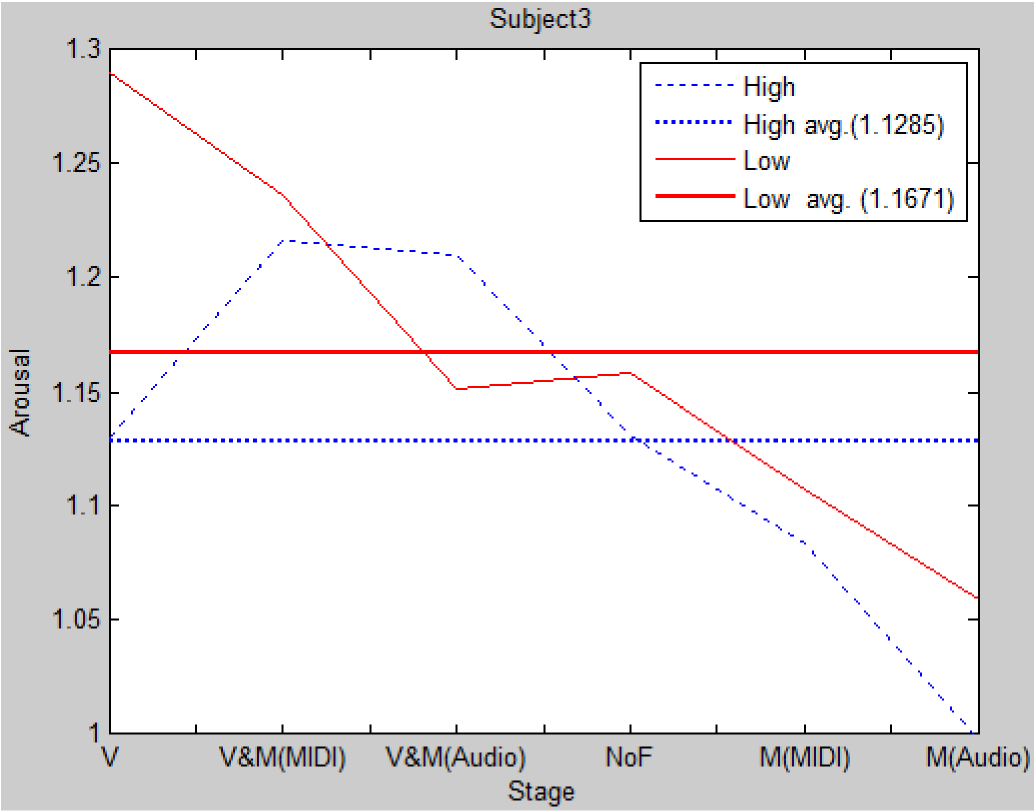
MIN				MAX
1	2	3	4	5

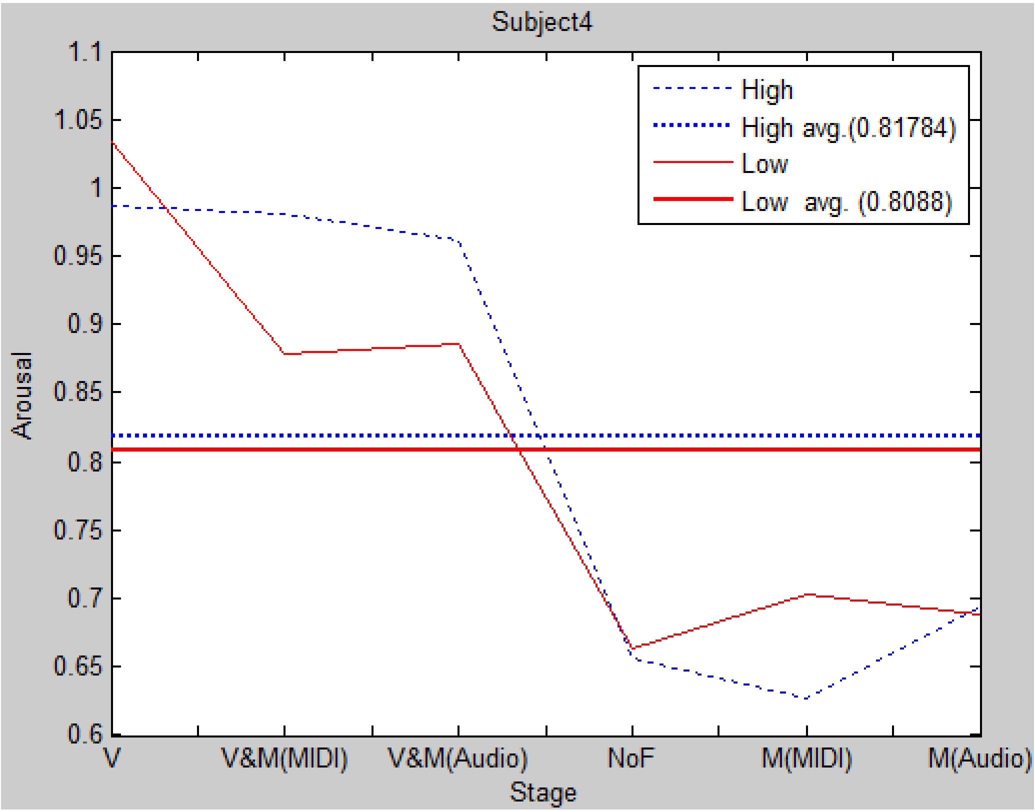
4. Please, mention something to improve the application.

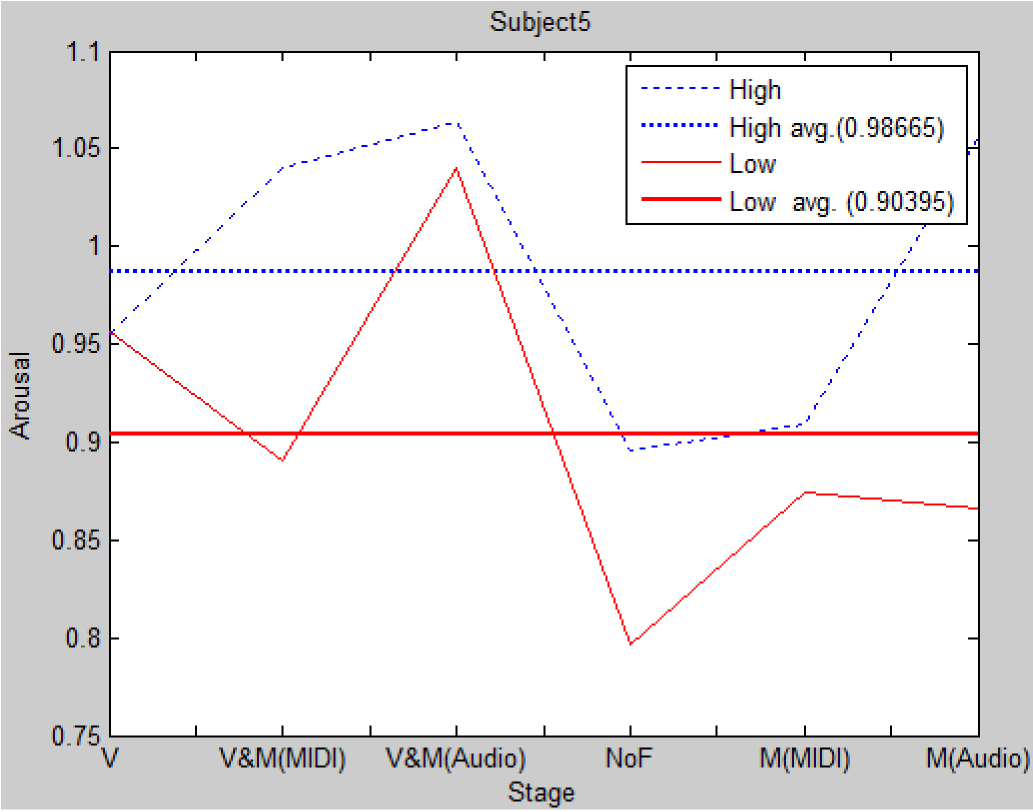
Appendix 5

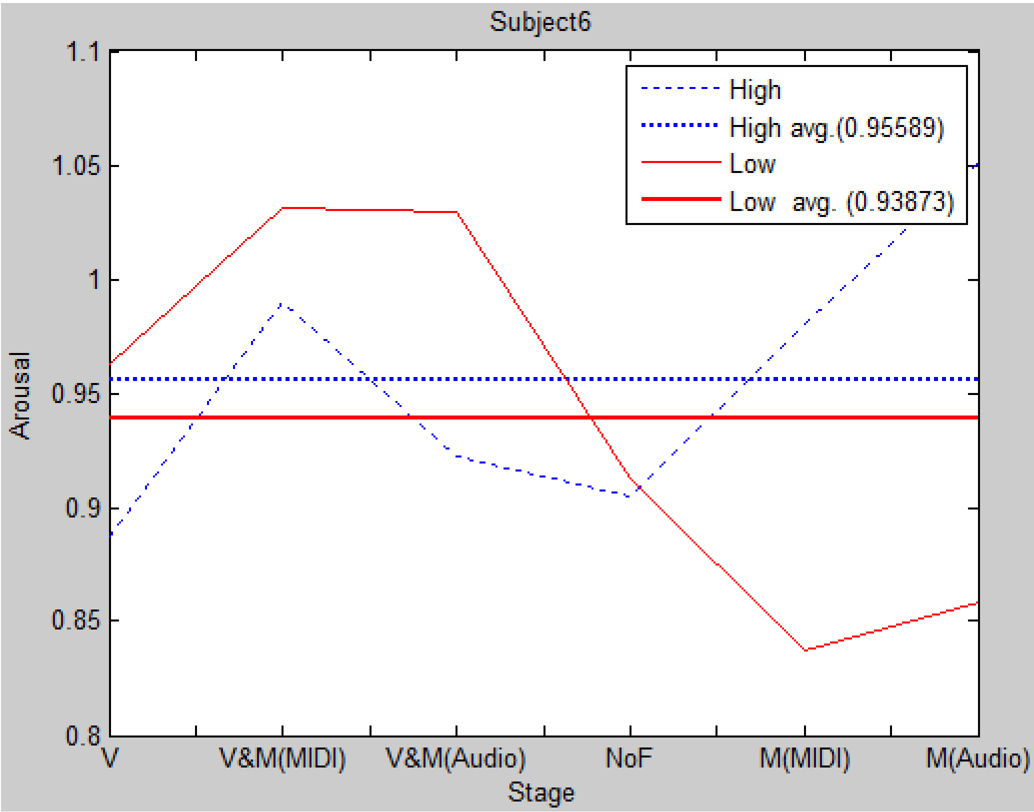

Subject1
High
High avg.(1.1455)
Low
Low avg. (1.1184)
Arousal
1.3
1.2
1.1
1
0.9
0.8
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)
Stage

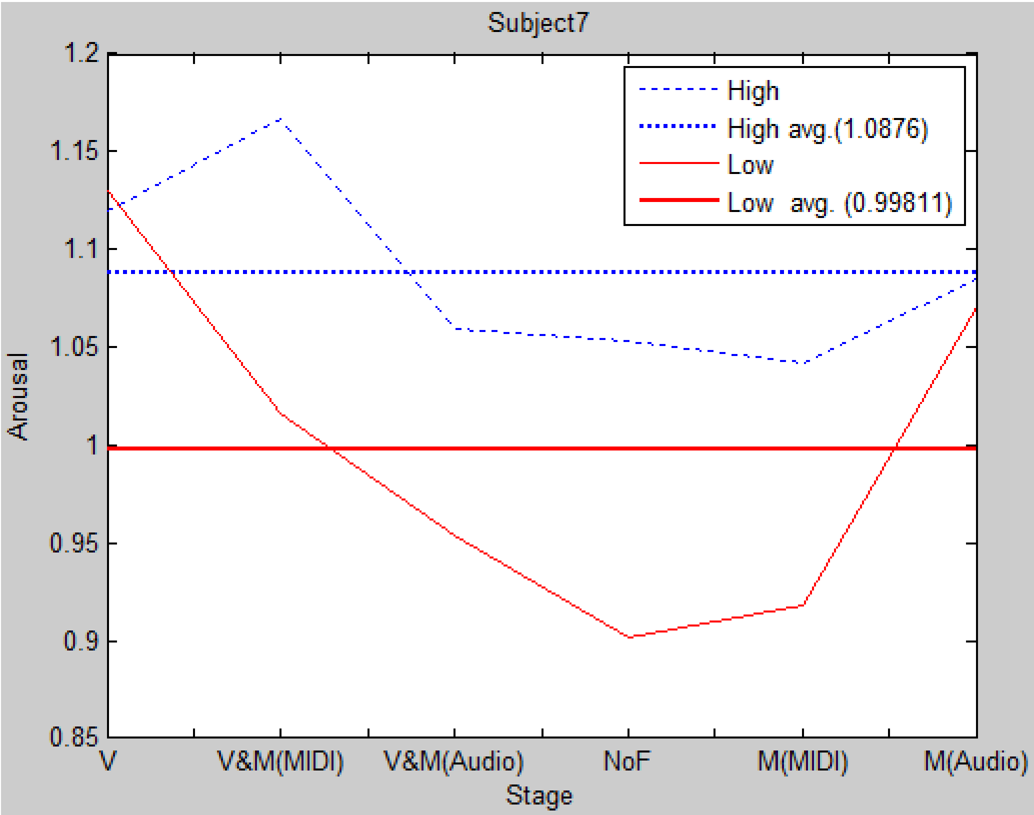

Subject2
High
High avg.(0.99585)
Low
Low avg. (0.97276)
Arousal
1.15
1.1
1.05
1
0.95
0.9
0.85
0.8
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)
Stage

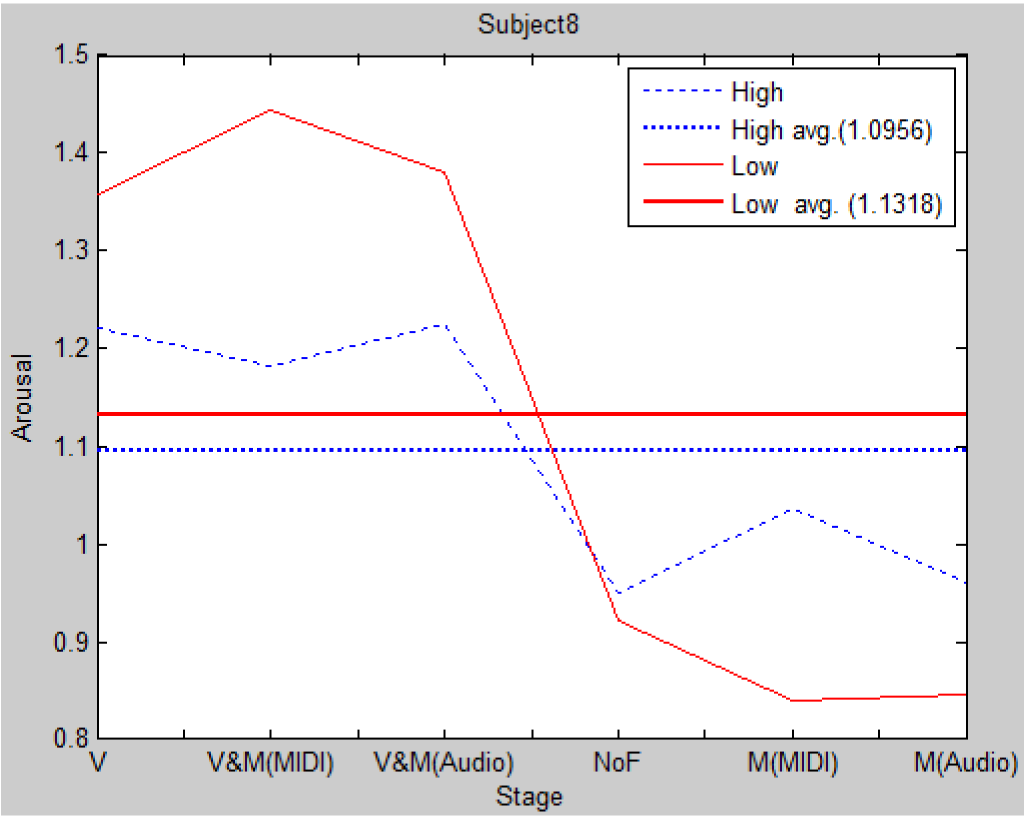

Subject3
Arousal
Stage
High
High avg.(1.1285)
Low
Low avg. (1.1671)
1.3
1.25
1.2
1.15
1.1
1.05
1
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)

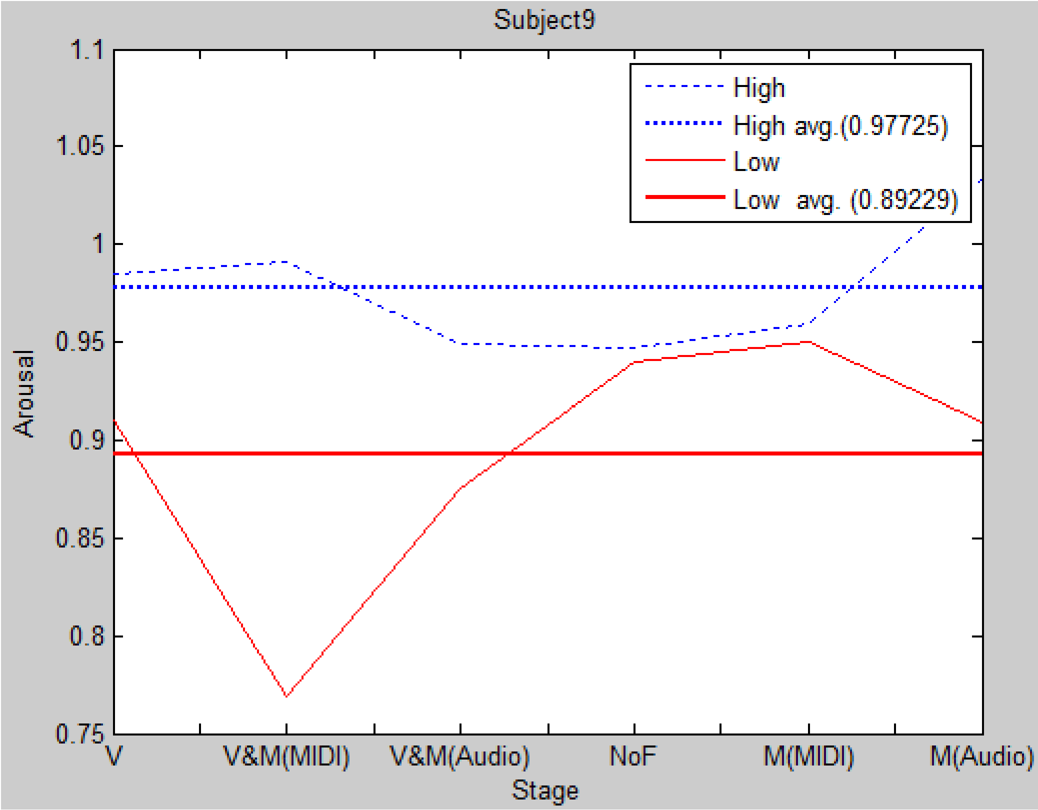

Subject4
Arousal
Stage
High
High avg.(0.81784)
Low
Low avg. (0.8088)
1.1
1.05
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)

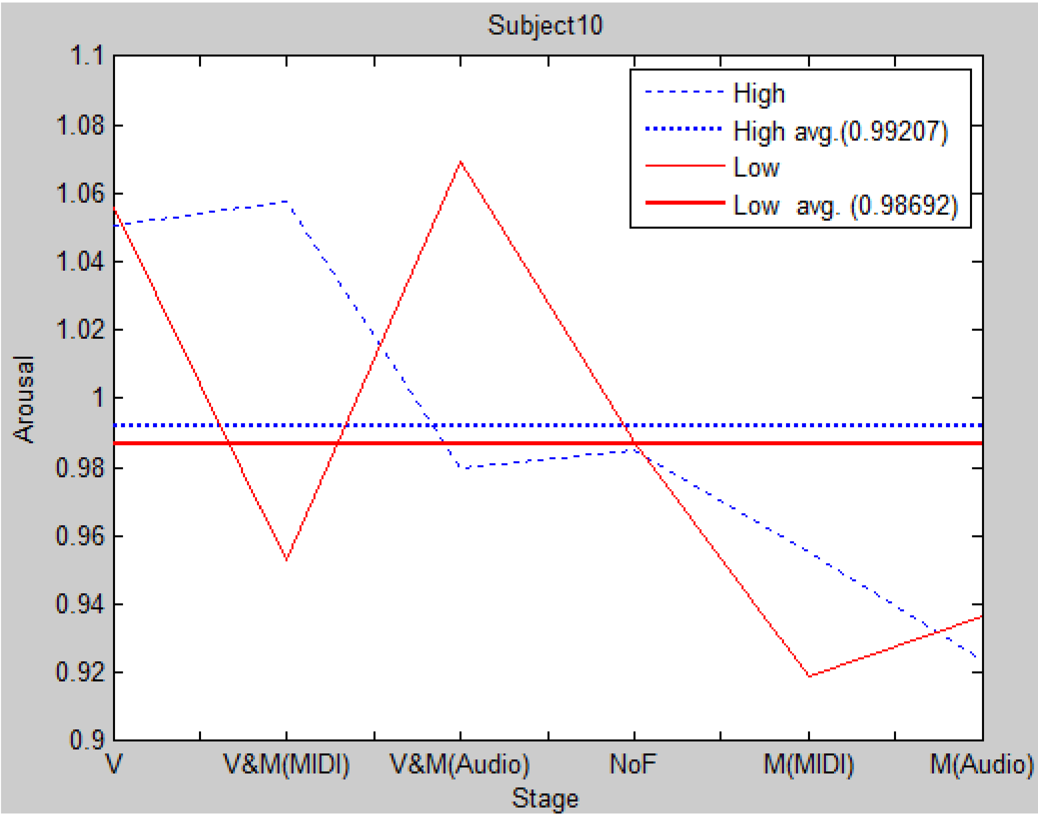

Subject5
1.1
1.05
1
0.95
0.9
0.85
0.8
0.75
Arousal
High
High avg.(0.98665)
Low
Low avg. (0.90395)
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)
Stage

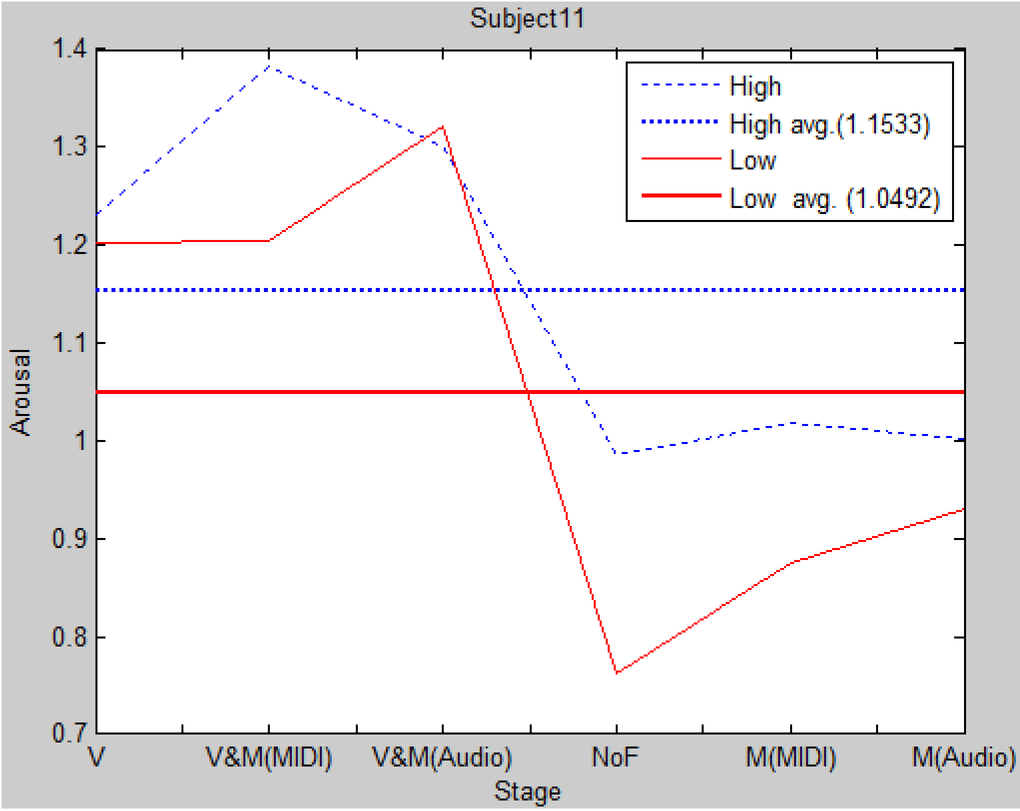

Subject6
1.1
1.05
1
0.95
0.9
0.85
0.8
Arousal
High
High avg.(0.95589)
Low
Low avg. (0.93873)
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)
Stage


Subject7
High
High avg.(1.0876)
Low
Low avg. (0.99811)
Arousal
1.2
1.15
1.1
1.05
1
0.95
0.9
0.85
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)
Stage


Subject8
High
High avg.(1.0956)
Low
Low avg. (1.1318)
Arousal
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)
Stage

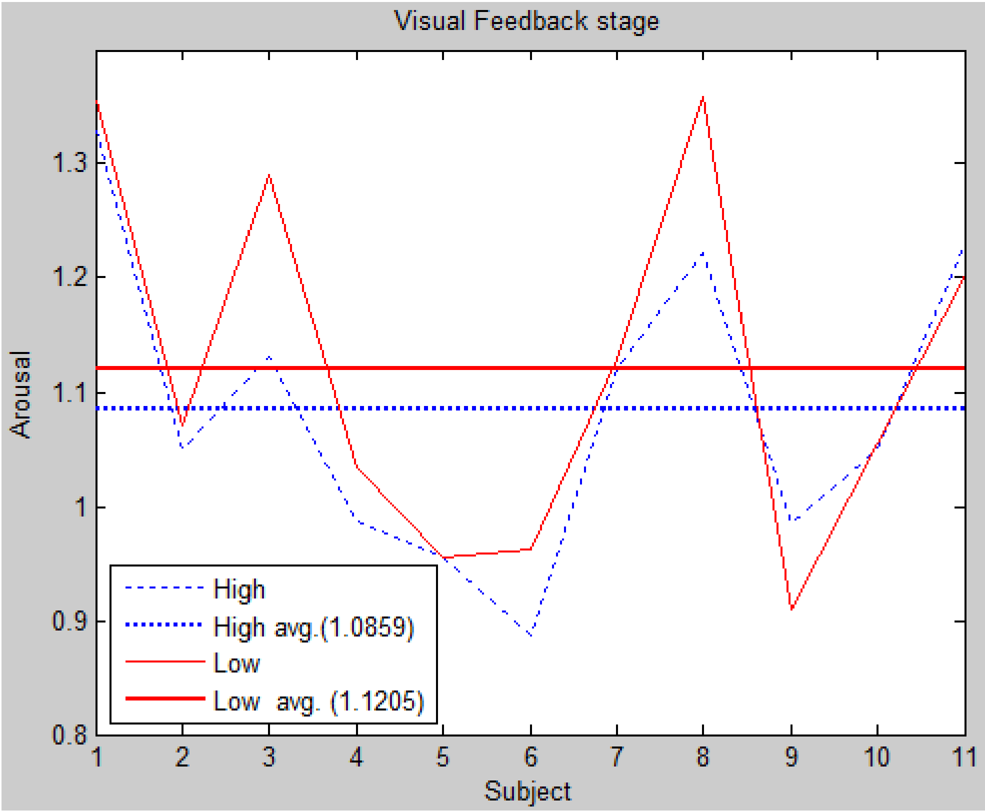

Subject9
High
High avg.(0.97725)
Low
Low avg. (0.89229)
Arousal
1.1
1.05
1
0.95
0.9
0.85
0.8
0.75
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)
Stage


Subject10
High
High avg.(0.99207)
Low
Low avg. (0.98692)
Arousal
1.1
1.08
1.06
1.04
1.02
1
0.98
0.96
0.94
0.92
0.9
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)
Stage

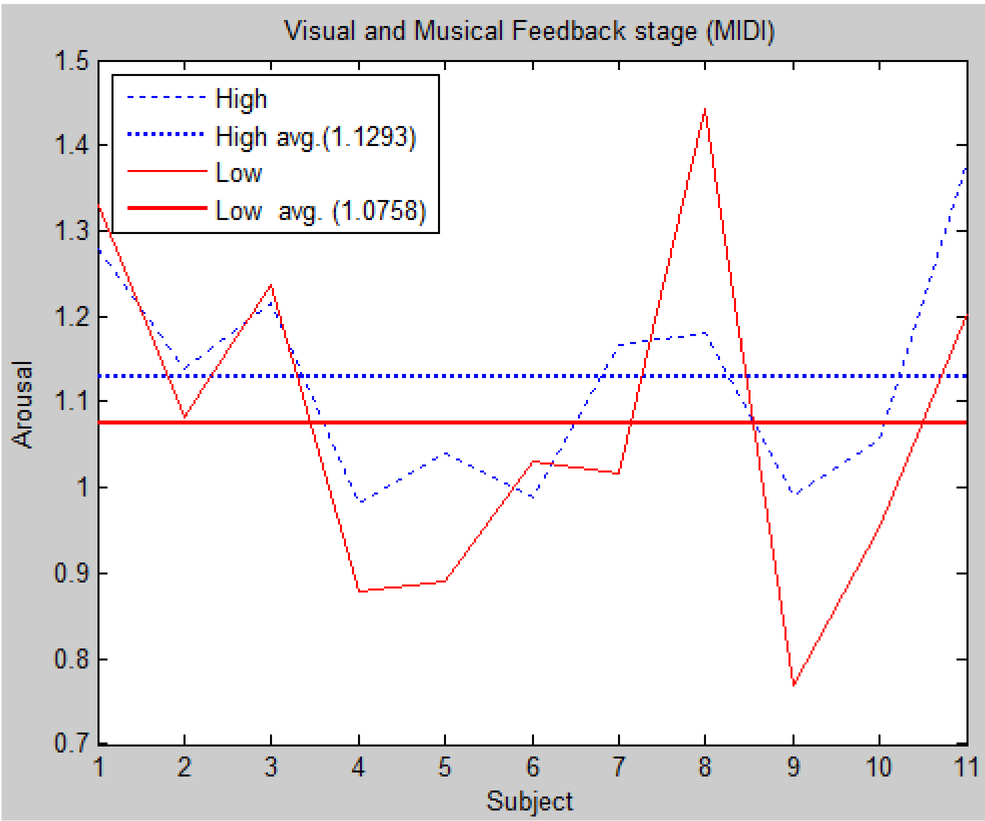


Subject11
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
Arousal
V
V&M(MIDI)
V&M(Audio)
NoF
M(MIDI)
M(Audio)
Stage
High
High avg.(1.1533)
Low
Low avg. (1.0492)

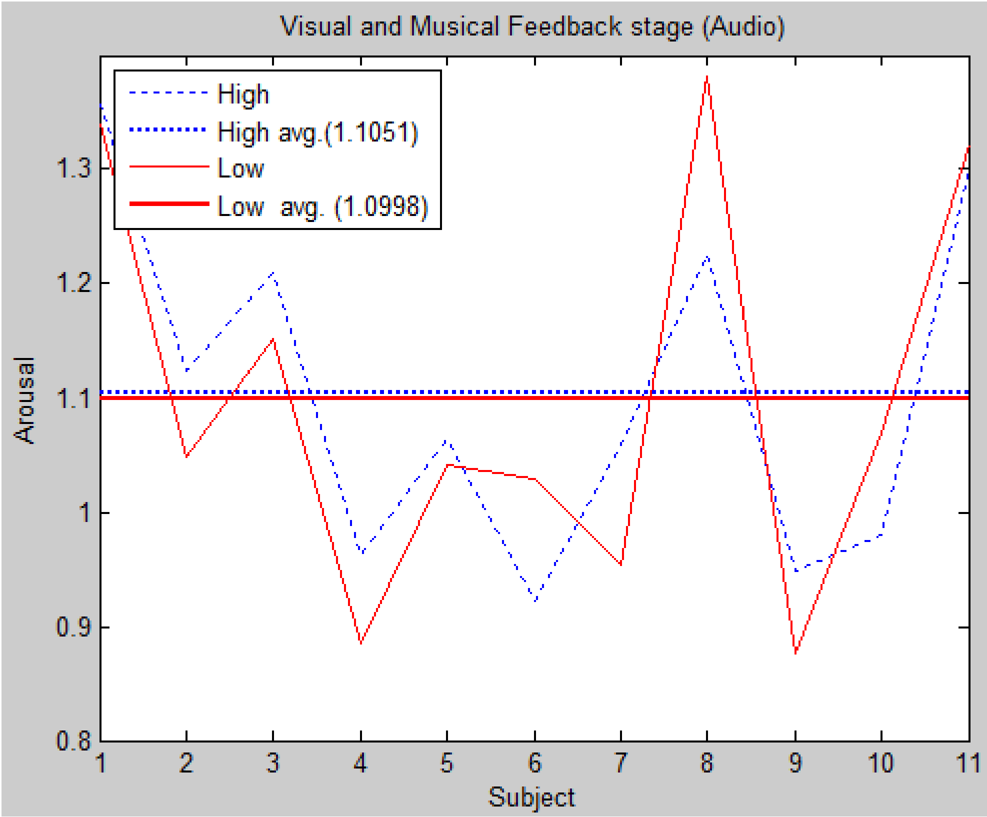
Appendix 6



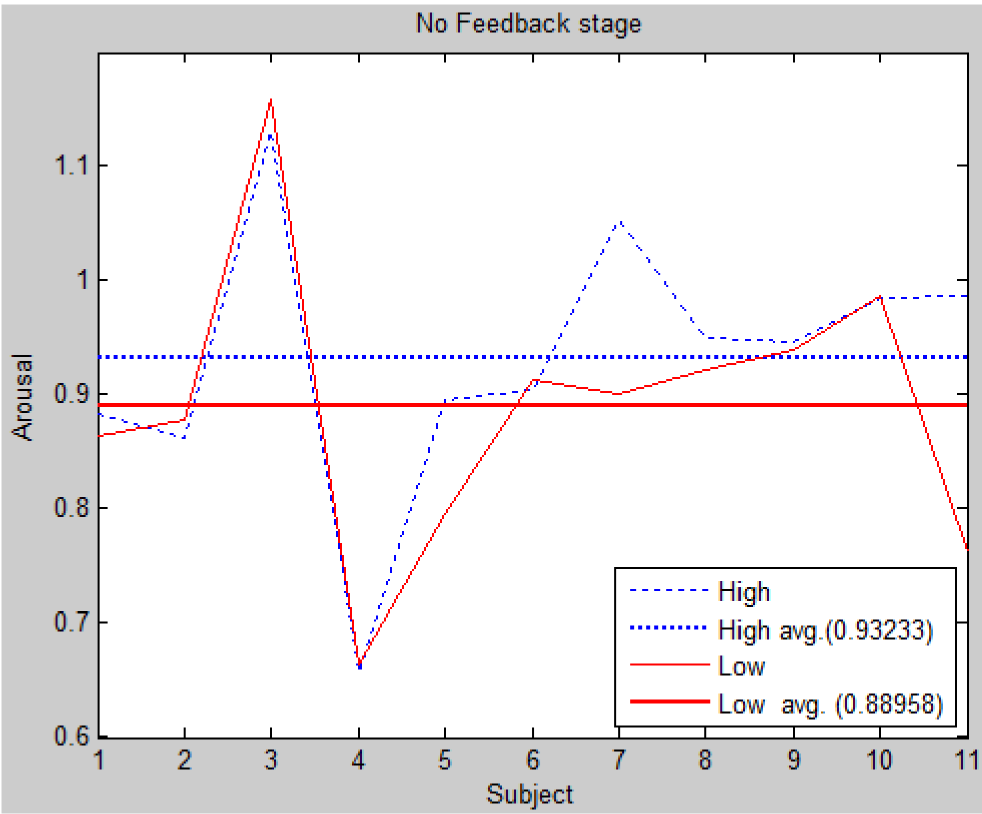
Visual Feedback stage
Arousal
Subject
High
High avg.(1.0859)
Low
Low avg. (1.1205)
0.8
0.9
1
1.1
1.2
1.3
1
2
3
4
5
6
7
8
9
10
11



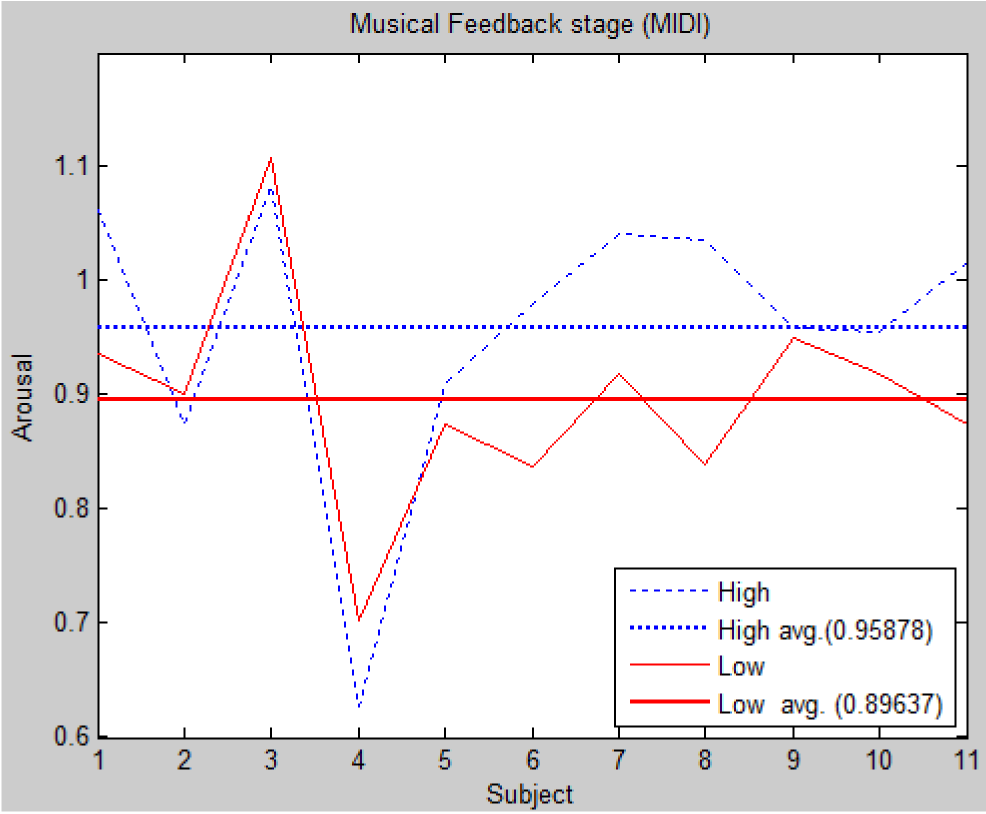
Visual and Musical Feedback stage (MIDI)
Arousal
Subject
High
High avg.(1.1293)
Low
Low avg. (1.0758)
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
1
2
3
4
5
6
7
8
9
10
11

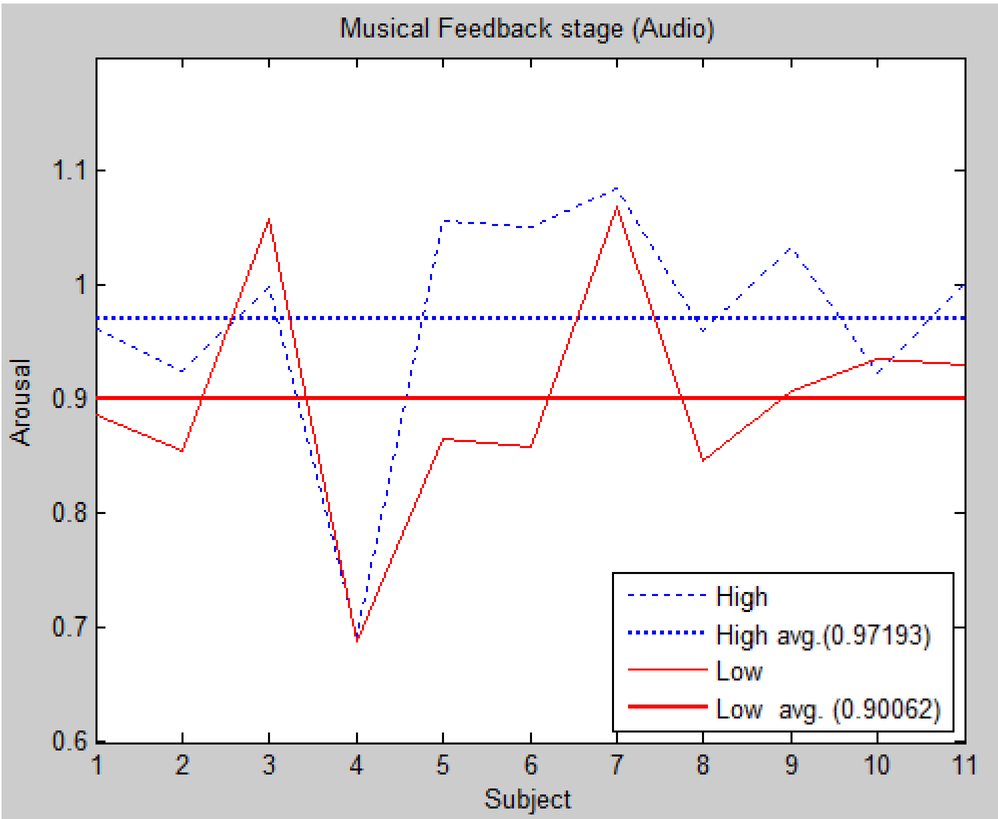


Visual and Musical Feedback stage (Audio)
High
High avg.(1.1051)
Low
Low avg. (1.0998)
Arousal
Subject
1.3
1.2
1.1
1
0.9
0.8
1
2
3
4
5
6
7
8
9
10
11



No Feedback stage
High
High avg.(0.93233)
Low
Low avg. (0.88958)
Arousal
Subject
1.1
1
0.9
0.8
0.7
0.6
1
2
3
4
5
6
7
8
9
10
11


Musical Feedback stage (MIDI)
Arousal
1.1
1
0.9
0.8
0.7
0.6
1
2
3
4
5
6
7
8
9
10
11
Subject
High
High avg.(0.95878)
Low
Low avg. (0.89637)


Musical Feedback stage (Audio)
Arousal
1.1
1
0.9
0.8
0.7
0.6
1
2
3
4
5
6
7
8
9
10
11
Subject
High
High avg.(0.97193)
Low
Low avg. (0.90062)