

Master thesis in Sound and Music Computing

Universitat Pompeu Fabra

Fusion of musical contents, brain activity and  
short term physiological signals for music-  
emotion recognition

Jimmy Jarjoura

Supervisor: Sergio Giraldo

Co-supervisor: Raphael Ramirez

July 2017



**Universitat  
Pompeu Fabra**  
*Barcelona*



## Acknowledgments

Foremost, I would like to express my sincere gratitude and appreciation to my advisor Sergio Giraldo for his continuous support during my study, for his patience, motivation, enthusiasm and immense knowledge. I am grateful to Dr. Rafael Ramirez and Batuhan Sayis from MTG for sharing their valuable experience and knowledge with me. I also would like to thank to Prof. Xavier Serra for giving me the opportunity to be part of this SMC master.

## Abstract

Detecting and classifying emotions from EEG signals has been reported to be a complex and subject dependent task. In this study we propose a multi-modal machine learning approach, combining EEG and Audio features for music emotion recognition using a categorical model of emotions. The dataset used consists of film music that was carefully created to induce strong emotions. Five emotion categories were adopted: Fear, Anger, Happy, Tender and Sad. EEG data was obtained from three male participants listening to the labeled music excerpts. The emotion classification accuracy from the stand-alone EEG system achieved 76.70%, 81.35% and 77.23% for three participants using Support Vector Machines (SVM).

EEG and Audio features were extracted, and later we applied machine learning techniques to study the improvement in the emotion recognition task using multi-modal features. Both feature types were extracted using a frame based approach.

Feature level fusion was adopted to combine EEG and Audio features. The results show that the multimodal system outperformed the EEG mono modal system with classification accuracies of 85.59%, 89.11% and 88.21% for three subjects.

Additionally, we evaluated the contribution of each audio feature in the classification performance of the multimodal system. Preliminary results indicate a significant contribution of individual audio features in the classification accuracy, we also found that various audio features that noticeably contributed in the classification accuracy were also reported in previous research studying the correlation between audio features and emotion ratings using the same dataset. These results conclude that there is some relevant and important acoustic information in the audio features which could improve the performance of the emotion recognition system.

Furthermore, we propose a framework for dealing with emotion recognition from physiological signals measured in short duration. Results show that certain features from skin conductance and heart rate variability were found efficient in the emotion classification task, thus the role of the activation of the autonomic nervous system in emotion recognition.

**Keywords:** Machine learning, EEG, multi-modal classification, audio features.

## Table of Contents

<b>1. Introduction</b>	<b>1</b>
Research questions	3
Goals	3
<b>2. State of the art</b>	<b>4</b>
<b>2.1. Mood classification from physiological information</b>	<b>4</b>
<b>2.2. EEG mood classification</b>	<b>6</b>
<b>2.3. Relationship between features and emotions</b>	<b>10</b>
2.3.1. Audio features and emotions	10
2.3.2. Emotions correlates with EEG and audio features	11
<b>3. Materials and methods</b>	<b>14</b>
<b>3.1. Dataset</b>	<b>14</b>
<b>3.2. Materials</b>	<b>14</b>
3.2.1. OpenVibe	14
3.2.2. Emotiv EPOC	15
3.2.3. EEGLab MATLAB	16
3.2.4. Bitalino and OpenSignals	16
3.2.5. Ledalab	18
<b>4. Method</b>	<b>19</b>
<b>4.1. Data collection and experiment procedure</b>	<b>19</b>
<b>4.2. EEG Feature extraction</b>	<b>19</b>
<b>4.3. Audio features extraction</b>	<b>21</b>
<b>4.4. Fusion of EEG and audio features</b>	<b>23</b>
<b>4.5. Skin conductance features extraction</b>	<b>23</b>
<b>4.6. Electrocardiogram (ECG) features</b>	<b>25</b>
<b>4.7. Fusion of EEG and skin conductance</b>	<b>26</b>
<b>4.8. Feature Classification</b>	<b>27</b>
4.8.1. EEG classification	27
4.8.2. Multimodal classification (EEG and Audio)	27
4.8.3. Multimodal classification (EEG and skin conductance)	28
<b>5. Results</b>	<b>29</b>
<b>5.1. EEG mood classification</b>	<b>29</b>
<b>5.2. Multimodal classification (EEG and Audio features)</b>	<b>32</b>
5.2.1. Multimodal classification accuracy	32
5.2.2. Learning curves	34
5.2.3. Contribution of individual Audio features	36
5.2.4. Comparison	39
<b>5.3. Multimodal classification (EEG and skin conductance)</b>	<b>40</b>
<b>5.4. ECG measurement</b>	<b>41</b>
<b>6. Discussion and conclusion</b>	<b>46</b>

## 1. Introduction

Music is a widely enjoyable experience that we rely on a daily basis. What makes music an important aspect in our daily life is the strong relation between music and emotions, as (Juslin & Laukka 2004) describes, there is an overwhelming evidence that musical events induce emotions in listeners.

Recent research in the field of Human Computer Interaction led the way for computers to understand, discern human emotions and perform various actions. (Picard 2003) states “Emotions play an essential role in rational decision-making, perception, learning, and a variety of functions”

Several researches have been conducted to recognize emotions using different modalities, for instance facial images, gestures, speech and physiological signals (Castellano et al. 2008); (Jerritta et al. 2011).

Monitoring ongoing brain activity electroencephalography (EEG), specifically, machine-learning approaches to characterize EEG dynamics associated with emotions gained increased attention. EEG emotion classification proved its potential in many applications such as musical ABCI (Makeig et al. 2011), neuromarketing (Lee et al. 2007), music therapy (Sourina et al. 2012) and implicit multimedia tagging (Aarts et al. 2014).

However, the EEG signals is known to be noisy and non-stationary (Daly et al. 2012), thus classifying emotions in music listening becomes a challenging task and can be greatly affected by considerable variation among different subjects. Moreover, the EEG device being used can have influence on the results and the classification performance, as for instance, some low cost headsets do not have a full brain coverage compared with more advanced EEG caps; The Emotiv EPOC headset (Emotiv systems, Inc.) has 14 electrodes compared with a 32 channels (Neuroscan, Inc.).

Early researches in music and emotion reported that certain properties of the musical structure induce certain emotions. Several musical characteristics has been studied and numerous features have been reported to be suggestive of discrete emotions such as mode, tempo and timber (Juslin & Laukka 2004);(Juslin & Sloboda 2001).

Additionally, there has been a rapid expansion of music information retrieval research toward automatic emotion recognition tasks to deal with the vast amount of accessible music digital data (Kim et al. 2010). Different acoustic features were investigated, developed and used in different machine learning approaches, and the performance of automated systems significantly improved in the last few years.

Moreover, the relationship between the audio features and emotional ratings by participants listening to music excerpts have been investigated (Laurier et al. 2009); (Trochidis et al. 2011); (Vempala & Russo 2012). The relationship explored confirm with the results from psychological studies.

When considering emotions in music, several models have been proposed. The discrete model of emotions and the dimensional model are the most dominant in music and emotion research (Juslin & Sloboda 2010);(Zentner & Eerola 2009). Although some neurological studies have reported the involvement of different processes in the two models (Gosselin et al. 2005);(Khalfa et al. 2008). However, some research concluded that there is no substantial differences and that there is high correspondence between the two models , except for a poorer resolution for the discrete model in identifying ambiguous examples (Eerola & Vuoskoski 2011).

From the aforementioned evidence, the noisy EEG signal, the limitation of the EEG caps and the significance of audio features in music emotion research, we hypothesize that a combination of brain activity and audio features would increase the emotion classification accuracy.

Furthermore, we investigate the combination of EEG and audio features while adopting a discrete model of emotions compared with the dimensional model which is widely used, using a film music dataset which was created to induce strong emotions. Moreover, the individual contribution of each audio feature in the classification task is evaluated and compared with other studies investigating the relationship between audio features and emotions.

Additionally, the fusion between EEG data and physiological is investigated as to assess the feasibility to include bio signals measured in a short duration.

## Research questions

From the fusion of different single modalities we pose the following research questions:

1. Are there significant differences between the performance of multimodal systems compared with single modalities in music emotion recognition?
2. Which individual audio features significantly improves the classification performance when combined with an EEG device with limited performance? Do the findings validate other research investigating the relationship between audio features and emotional ratings from participants?
3. Is it possible to combine EEG signals with physiological signals measured in a short duration?

## Goals

This thesis aims to achieve different goals:

- Evaluate the EEG emotion classification performance and the contribution of audio features.
- Investigate which audio features has perceptual significance to be incorporated in a multimodal scheme.
- Evaluate learning curves and training size variations for multimodal schemes.
- Evaluate the performance of different machine learning algorithms and feature selection methods.
- Propose a framework for the fusion of EEG and short term physiological signals for music emotion recognition.

## 2. State of the art

### 2.1. Mood classification from physiological information

The recent technological advancements in human-computer interaction enabled the study of human emotions through the analysis of bio signals and physiological information. Human emotions can be recognized by several approaches such as gesture, facial images, physiological signals and neuroimaging techniques.

Numerous findings in psychophysiology reported that the activation of the autonomic nervous system changes when emotions are elicited (Levenson 1988). Bio signals such as Galvanic Skin Response (GSR) and Electrocardiogram (ECG) provide an *invisible* proof of affective arousal compared to visible modalities such facial expression or body gestures (Gunes & Pantic 2010).

Galvanic Skin Response provides a measure of the skin conductance (SC). GSR (Galvanic Skin Response) is popularly called EDA (Electrodermal Activity). EDA refers to the variation of the electrical properties of the skin in response to sweat secretion. EDA is reported to increase as a response to stress or overall arousal.

Electrocardiogram (ECG), measure the activity of the heart, it can be used to measure heart rate and inter-beats intervals to determine the heart rate variability (HRV). A low heart rate variability indicates a relaxed state, whereas a higher rate variability indicates a state of stress.

Hence both EDA and HRV allow to monitor the Autonomic Nervous System (ANS).

(Kim & André 2008) used four-channel biosensors to measure electromyogram, electrocardiogram, skin conductivity, and respiration changes to recognize emotions from three participants during many weeks. The songs were handpicked by the participants. Using the 2D emotion model with 4 emotions (Joy, anger, sad and pleasure), and investigating several features, they showed that the best accuracy was obtained using a novel emotion-specific multilevel dichotomous classification (EMDC) method with an accuracy of 95 percent and 70 percent for subject-dependent and subject-independent classification, respectively.

(Zong & Chetouani 2009) proposed a feature extraction technique based on the Hilbert-Huang

Transform (HHT) method for emotion recognition from physiological signals. Four kinds of physiological signals were used for analysis: electrocardiogram (ECG), electromyogram (EMG), skin conductivity (SC) and respiration changes (RSP). They used 25 recordings of two minutes length from four emotions (joy, anger, sadness and pleasure). They achieved 76% accuracy using SVM. Over the last decade, research on emotion recognition using physiological signals has been widely studied. Even though, the physiological information exhibits subjectivity, sensitivity to movement artifact and a need to process the data and extract features, however it provides an opportunity to recognize affect changes that cannot be easily perceived visually (Jerritta et al. 2011). The table below summarizes the details of the work that has been done so far in classifying the various emotions using physiological signals from music stimuli.

*Table 1: Overview of research dealing with emotion prediction from physiological signals*

Authors	Bio signals used	No of subjects	Emotions	Stimuli used	Feature extraction	Classification	% accuracy
(Kim & André 2008)	Electromyogram Electrocardiogram Skin Conductance Respiration	3	Joy Anger Sad Pleasure	Music	Statistical and Energy based features-Sub band Spectrum, Entropy	EMDC and Linear Discriminant Analysis	95(User Dependent) 70(User Independent)
(Zong & Chetouani 2009)	Electrocardiogram Electromyogram Skin Conductance Respiration	MIT database	Joy Anger Sad Pleasure	Music	Hilbert Huang Transform (Fission and Fusion)	Support vector machine	76 (Fission, User Dependent) and 62 (Fusion, User Dependent)
(Cheng & Liu 2008)	Electromyogram	MIT database	Joy Anger Sad Pleasure	Music	Daubechies5 Wavelet transform	Neural Network	82.29 (User Dependent)
(Zhu 2010)	Electromyogram	MIT database	Joy Anger Sad Pleasure	Music	Six scale Daubechies Wavelet Transform	Support vector machine	83.30 (User Dependent)
(Hönig et al. 2009)	Electrocardiogram Electromyogram Skin Conductance Respiration	MIT database	Joy Anger Sad Pleasure	Music	Moving features and sliding features (Recursive)	Linear Discriminant Analysis	83.4 (User Dependent)

### 2.1.1. Window size

It is important to discuss the duration of emotion in the experiments. Determining the length of the temporal window for automatic affect analysis depends on the modality and the target emotion (Gunes & Pantic 2010). (Levenson 1988) reports that the overall duration of emotions approximately falls between 0.5 and 4 seconds. However, the different literature investigating physiological signals in terms of emotion recognition do not report the best window size to be adopted. The research papers usually adopt different temporal windows. For example (Kim 2007) used short window sizes (2-6 seconds for speech, 3-15 seconds for bio signals) in contrast with other literature using larger durations (160 seconds (Kim & André 2008)).

## 2.2. EEG mood classification

In 2000 Choppin used EEG signals to classify six emotions based on a dimensional representation (valence and arousal) (Choppin 2000). He summarized that positive emotions are characterized by a high frontal coherence in alpha, and high right parietal beta power. Excitation is characterized by higher beta power and coherence in the parietal lobe, plus lower alpha activity. The strength of an emotional feeling is characterized by an increase in the beta/alpha activity ratio plus an increase in beta activity at the parietal lobe.

In 2003 Ishino and Hagiwara (Ishino & Hagiwara 2003) proposed a system for feeling estimation. They adopted four feeling states (joy, anger, sorrow, and relaxation). They used neural networks to categorize emotional states based on EEG features. The obtained accuracy ranges from 54.5% to 67.7% for each of four emotional states.

In 2004 Takahashi (Takahashi 2004) proposed a multimodal emotion-recognition system based on bio-potential signals (EEG, pulse and skin conductance). For EEG measurements, Takahashi used a headband of three dry electrodes to classify five emotions (joy, anger, sadness, fear, and relaxation). Using support vector machine (SVM), the result showed that the recognition rate reached an accuracy of 41.7 % for five emotions.

In 2006 Oude (Bos 2006) used the BraInquiry EEG PET device to recognize emotions. She uses a limited number of electrodes and trains a linear classifier based on Fishers discriminant analysis. She considers audio, visual and audiovisual stimuli and trained classifiers for positive/negative, aroused/calm and audio/visual/audiovisual.

In 2007 Heraz et al. (Heraz et al. 2007) established an agent to predict emotional states during listening. Eight emotional states were adopted, and k-nearest neighbors classifier showed the best accuracy around 82.27%.

In 2009 Chanel et al. (Chanel et al. 2009) used EEG time-frequency information as features, and three emotional states, and with SVM classifier they achieved a 63% average accuracy.

In 2009 Zhang and Lee (Zhang & Lee 2009) proposed a system to categorize emotions reflected by natural scenes and image viewing. Functional magnetic resonance imaging (fMRI) and EEG were used to analyze the natural scene's effect on human emotions. The system employed asymmetrical characteristics at the frontal lobe as features and SVM as a classifier. They achieved an accuracy of  $73.0\% \pm 0.33\%$ .

In 2010 Lin et al. (Lin et al. 2010) used four emotional states: joy, anger, sadness, and pleasure, and adopted machine learning techniques to categorize EEG signals according to subject self-reported emotional states during music listening. Subjects listened to 16 30-s Oscar's film soundtracks. EEG features were extracted using short-time Fourier Transform (STFT) with a nonoverlapped Hanning window of 1 s applied to each of 30 channels of the EEG data to compute the spectrogram. Spectral power features were extracted from 30 electrodes in addition to 12 asymmetry indexes from electrode pairs. The results show that that the differential asymmetry of hemispheric EEG power spectra provided the best classification accuracy compared with spectral power of the electrodes which exhibits a larger number of features. Using SVM, the best classification accuracy for asymmetry indexes is 82.29% compared with 71.15% for spectral power of 30 electrodes.

In 2012 Ramirez and Vamvakousis (Ramirez & Vamvakousis 2012) used the Emotiv EPOC device to detect emotion from EEG signals. They characterized emotional states by computing arousal and valence levels as the prefrontal cortex beta to alpha ratio and as the alpha asymmetry between lobes respectively. The experiments included selected sounds from IADS library of emotion-annotated sounds (Lang et al. 1999). The twelve sound stimuli chosen are situated in the extremes of the dimensional plane (three in each plane).

The EEG signal is measured in four locations in the prefrontal cortex: AF3, AF4, F3 and F4. They used beta/alpha ratio to characterize the arousal state. To classify emotional states, they applied Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). For the high-versus-low arousal, and the positive- versus-negative valence classifiers the average accuracies they obtained for SVM with radial basis function kernel classifier were 77.82%, and 80.11%, respectively.

### 2.2.1. Window size

Different literature investigating emotion detection from EEG signals do not specify the optimal windows size to adopt. Various window sizes were used in previous research; (Yurci 2014) adopted 1 second windows size and 1 second hop size. (Eaton et al. 2014) adopted the last 10 seconds of 30 windows (which is equal to the length of each music clip). (Choppin 2000) used different window sizes depending on the type of experiment (5.12 seconds and 10.24 seconds). (Shin et al. 2014) used 8 seconds of EEG signals for their mood-based music recommendation system. (Lin et al. 2010) used Fourier transform (STFT) with a non-overlapped Hanning window of 1s was applied to each of 30 channels of the EEG data to compute the spectrogram. For (Sawata et al. 2015) in their paper, the time length of an EEG segment and an overlapping are 1.0 and 0.5 sec, same as audio features. (Müller et al. 2008) used a 2 seconds window for their offline EEG signal analysis. In (Lin et al. 2007), EEG data were epoched with a Hanning window of 1 sec width (500 points) and 50% overlap. (Wang et al. 2014) in their emotional state classification from EEG data during movie induction experiment, they compared the classification performance using power spectrum across all EEG frequency bands with four different length time windows of 0.5 s, 1 s, 1.5 s, and 2 s. They adopted 1 second as the time window length.

Table 2: Overview of research dealing with emotion classification from brain activity (EEG)

Author	Device	Stimuli type	No of emotions	Features	Algorithm	Accuracy
<b>Ramirez, Vamvakousis 2012</b>	EPOC	Audio stimuli	4 emotions	Arousal/valence Beta/Alpha	LDA and SVM	77.82%, and 80.11%,
<b>Chopin 2000</b>	EEG (13 electrodes)	Pictures/sound	6 emotions	Arousal/valence Beta/Alpha	Neural networks	64%
<b>Oude 2006</b>	BraInquiry EEG PET	audio/visual/audiovisual.		Modality, Arousal and Valance Levels	Fishers discriminant analysis. and VA-ARO	~80%
<b>Takachi 2004</b>	EEG,pulse,skin conductance (three electrodes)	Audio-visual	5 emotions	Alpha,beta,theta,delta +skin +pulse	SVM	41.7%
<b>Lin 2010</b>	EEG signals according to subject self-reported emotional states	Music listening	4 emotions	Spectral power of the electrodes and asymmetry indexes	SVM	~82%
<b>Liu Y.et al /2012</b>	Emotiv EPOC (14 electrodes)	audio/visual/audiovisual	6 emotions		Fractal dimensions and VA-ARO	
<b>Heraz et al./2007</b>	Pendant EEG (3 electrodes)	IAPS (International Emotional Picture System)	8 emotions	Main amplitudes of brainwaves in different frequency bands	k-NN	~80%
<b>Ramirez R.et al/ 2012</b>	Emotiv EPOC ( 4 electrodes-frontal cortex)	IADS	Arousal and valence Levels	Arousal and Valence features	LDA and SVM	77.82 % 80.11 %

## 2.3. Relationship between features and emotions

### 2.3.1. Audio features and emotions

In 2009 (Laurier et al. 2009) explored the relationship between audio features and emotions in music. They adopted a dataset consisting of 110 excerpts from film soundtracks that was evaluated by 116 listeners. The dataset is annotated with 5 basic emotions (fear, anger, happiness, sadness, tenderness) on a 7 points scale. The dataset was created in a study by (Eerola & Vuoskoski 2011). The main advantage in using soundtracks is that they were initially composed to convey emotions. The concluded results are aligned with previous studies on emotions perception. A positive correlation was found between dissonance and fear and anger in addition to a negative correlation with sad and tenderness. From music mode they showed that major modes are more correlated to positive emotions whereas minor modes are more correlated with negative emotions. Onset rate was found correlated with the happy category whereas loudness was found correlated with happy and anger.

(Trochidis et al. 2011) investigated the relationship between musical audio features and perceived emotions in Contemporary Western music. They adopted a dimensional approach for emotional labelling with 27 excerpts selected by music theorists and psychologists. The results show that low level spectral and temporal features such as flatness and chroma are efficient in modeling the arousal dimension, whereas high-level features such articulation, pulse clarity, mode and brightness succeed to measure the valence dimension. They conclude their study with the importance of using physiological signals such as EEGs “to bridge the semantic gap between high level knowledge related to the cognitive aspects of emotion and low level acoustical features”.

(Vempala & Russo 2012) adopted neural networks for predicting arousal/valence ratings from participants listening to 12 classical music excerpts from 12 classical composers. They extracted 13 low and mid-level features related to the dynamics, rhythm, timbre, pitch and tonality. Concerning the correlation between audio features and emotion rating, they found strong positive correlations between arousal and five audio features: pulse clarity, zero crossing rate, spectral

centroid, spectral roll-off and brightness. Low energy and mode we found positively correlated with valence.

### 2.3.2. Emotions correlates with EEG and audio features

Given that music is a highly complex phenomenon, investigating emotions induced by music should consider different aspects of emotion induction. Considering the subjectivity of emotion among listeners, it is advantageous to adopt physiological measurements as correlates of emotional responses. (Juslin & Laukka, 2004) reported that emotion is inferred from three kind of evidence: (a) self-report, (b) behavior and (c) physiological reactions such as EEG, ECG, blood pressure etc. Furthermore, they extend their findings by claiming that “physiological measures should necessarily be used in connection with other measures (such as self-report)”.

(Daly et al. 2015) combined both physiological measurements and acoustic properties to effectively predict emotional responses to a piece of music given that EEG is a noisy non stationary signal (Daly et al. 2012). Thirty-one individuals participated in this study and listened to 40 excerpts of 12s of film soundtracks created by (Eerola & Vuoskoski 2011).

Different acoustic features were extracted from the audio data: temporal features, spectral features, perceptual features, cepstral features, and features describing the beat of the music.

Concerning EEG features, they extracted band-power features and pre-frontal asymmetry features. A feature selection method based on principal component analysis was adopted to select a subset of features from both the audio and EEG. A linear regression model was used with each response PC to predict the emotion responses of the participants (for a review: (Daly et al. 2014);(Daly et al. 2015)). Their results showed that when combining EEG and acoustic features, they were able to predict emotional response more accurately than when using either of the two features alone. They found a subset of features that allow to train a linear regression model along each of the first 3 PCs that correspond to each of the three dimensional axes of Schimmack and Grob model of affective states (valence, energy-arousal, and tension-arousal) (Schimmack & Grob 2000).

More specifically, they found that features related to the variance of Mel cepstral coefficients in high frequency bands was related to valence. Low frequency Mel cepstral coefficient was found related to Energy-arousal and Tension-arousal.

(Yurci 2014) in his master thesis, studied the correlation between EEG signal characteristics and three emotional states: happy, sad and relax. Furthermore, he studied the correlation between musical features and EEG signals. Six songs with emotional content were selected. Three of the songs were retrieved from a dataset created in (Laurier 2011). The other three songs were selected by each participants depending on their own preference. Eleven subjects participated in this study, and EEG signals were measured using the Emotiv EPOC device. Additionally, subjects rated their emotional states in relevance to which degree they felt the emotional label of the music stimulus. For each electrode their  $\theta$  (theta),  $\alpha$  (alpha) and  $\beta$  (beta) bands were calculated. Additionally, high-level EEG features (arousal and valence variations) were computed. In total 83 EEG features were extracted:  $\theta$  (theta),  $\alpha$  (alpha) and  $\beta$  (beta) bands, high-level EEG features and raw EEG signals.

Concerning EEG classification, the highest result was achieved by the K-NN algorithm which was 98.2%. In terms of musical and EEG features correlation, for the subject's selected songs,  $\beta$  T7 EEG feature had the highest correlations. Musical features that were correlated with  $\beta$  T7 EEG feature were MFCC (1), Entropy, Spectral Flux, Spectral Centroid, MFCC (2), MFCC (3) and Zero-crossing Rate which were belonging to timbral properties of the music while Rms Slope, STD (Rms), Rms were related with dynamic properties of the music. The average correlation coefficient was greater than 0.28.

On the other hand, the correlation results for the common songs were different. The Valence Raw FC6 showed the bigger number of correlations (8 correlations). Overall, the common songs scenario showed more correlations compared to the scenario with subject's selected songs.

(Sawata et al. 2015) proposed a system that uses EEG signals to calculates new audio features that are suitable to represent a user's music preference. The method uses Canonical Correlation Analysis (CCA) on audio features and EEG signals measured when the user listens to a favorite musical piece. This projection enables the identification of EEG-audio subsets that show the best correlation. They extracted different EEG features such as Zero Crossing Rate, Mean Frequency, Spectral Entropy, and other features related to different wave bands (alpha, beta, theta and gamma). Furthermore, they extracted different spectral, timbre, tonal, and rhythmic audio features. The kernel CCA that uses a projection to calculate new EEG-audio features reflects a user's music

preference and extracts linear and non-linear correlations. They used a dataset of 60 musical pieces with a duration of 15s from 12 genres. The participants rated their liking of a music piece on a scale from 1 to 5. In total 5 subjects participated in the study, and EEG signals were extracted from 12 channels. For the classification approach, SVM classifier was adopted. The results show that the CCA based audio features did not outperform the one using only user's EEG features, however the methods based on kernel CCA i.e., the new EEG-based audio features outperforms all comparative methods.

(Thammasan et al. 2016) presented a study of multimodality using EEG and musical features extracted from synthesized MIDI pop songs. Fractal Dimension (FD) approach was adopted to extract EEG features from 12 electrodes in addition to five asymmetry indexes of five right-left electrode pairs. MIDI files were converted to WAV files and several musical features related to dynamics, rhythm, timbre and tonality were extracted. A decision level fusion was adopted in which each classification modality is processed independently and the output of the classifier are combined to yield the final result depending on a weight factor that determines the degree of contribution of each modality. A binary emotion classification for Arousal/Valence was investigated. The results show that music mono-modality achieved the best performance regardless of the window sized adopted. The fusion of EEG and audio features outperformed other modalities. However, increasing the contribution of EEG features in the multimodality decreased the classification accuracy. Lastly the class imbalance and the limited number of songs in the dataset they adopted was discussed.

## 3. Materials and methods

### 3.1. Dataset

The dataset adopted was created by (Eerola & Vuoskoski 2011). This dataset consists of film music that is created with the intention to induce strong emotions and could be regarded as a “neutral” stimulus in terms of genre, preferences and familiarity. Additionally, unfamiliar excerpts were chosen as to eliminate as much as possible the effect of *episodic memories*. To construct a large selection of film music, 12 expert musicologist selected 360 audio clips (equally representative of both discrete emotions and three-dimensional model) in relevance with different criteria. The duration of the tracks is between 10 and 30 seconds.

To form a refined set of musical examples, and with relevance to different criteria, 110 film soundtracks excerpts which contains tracks representative of both discrete and categorical models were chosen. Moreover, a large pilot study was conducted, in which 116 university students rated the refined dataset based on the 3D and discrete models.

The discrete model consisted of five emotions: Anger, fear, happy, tender and sad, whereas the categorical model consists of excerpts representative of the six extremes of three bipolar axes (Valence, energy arousal and tension arousal). The dataset does not only contain the best examples of the targeted emotion, but also tracks with moderate representation of a particular emotion.

For this thesis, 8 excerpts from each discrete emotion were adopted. 4 excerpts with the highest ratings in the targeted emotion, and 4 excerpts with moderate rating, which constitutes 40 excerpts equally distributed among 5 discrete emotions (Anger, fear, happy, tender and sad).

### 3.2. Materials

#### 3.2.1. OpenVibe

OpenVibe is an open-source software meant to create Brain-Computer Interaction (BCI) applications. For the purpose of the thesis, a Lua code was developed which receives a key input from the computer’s keyboard, and depending on the key pressed, the code will start playing a

track from the 5 emotion categories (Fear, angry, happy, sad and tender). At the exact time when the key is pressed, the Lua simulator in OpenVibe will mark the beginning of the brain activity measurement. Additionally, depending on the duration specified, the code will mark the end of the measurement. Furthermore, the brain signals are stored in GDF files for further processing at later stages.

### 3.2.2. Emotiv EPOC

EEG data was collected by the Emotiv EPOC neuroheadset (Emotiv Systems Inc.) which has 14 electrodes, plus 2 electrodes for reference and noise reduction. The electrodes are located and labeled according to the international 10-20 system (Niedermeyer & Silva 2004). Following the international standard, the available locations are: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4. Figure 1 shows the 14 Emotiv EPOC headset electrode positions. The EEG signals are transmitted wirelessly to a laptop computer where OpenVibe is running.

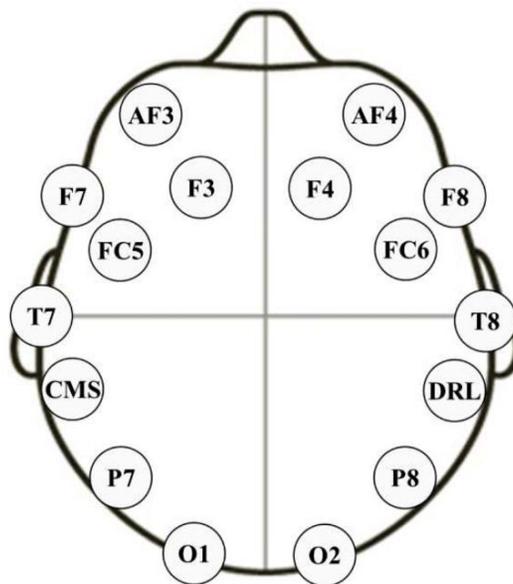


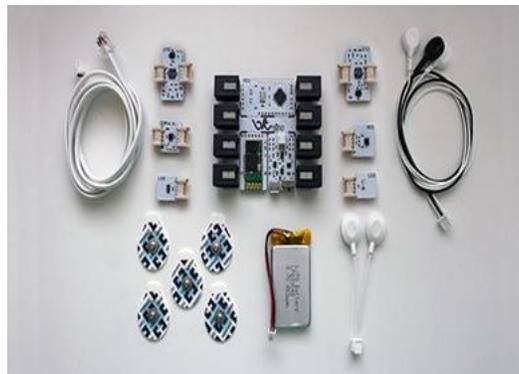
Figure 1: Available electrode positions on the Emotiv EPOC according to the 10-20 electrode placement system

### 3.2.3. EEGLab MATLAB

EEGLAB is an interactive Matlab toolbox for processing continuous and event-related EEG, MEG and other electrophysiological data (Delorme & Makeig 2004). Using EEGLab functions, a MATLAB code was developed in order to automatically process the input brain signals, extract the correspondent features, and outputs a CSV file ready to be incorporated in machine learning approaches. The developed code includes automatically extracting all epochs related to the 5 discrete emotions (Fear, angry, happy, sad and tender), depending on the event number specified by EEGlab for each of the 5 emotions which are initially marked by the input stimuli in OpenVibe. Then, the code preprocesses the input brain signals and extracts different features from different frequency bands. The detailed feature computation is discussed in the next section. The different features are computed based on window and hop sizes specified by the user. Finally, each frame computed is labeled with the corresponding discrete emotion tag, and the output file includes the different frames with their correspondent features and their emotion tags.

### 3.2.4. Bitalino and OpenSignals

Bitalino (Guerreiro et al. 2013) is a low cost hardware platform for bio signals acquisition and wireless transmission in real time. OpenSignals software was used to test the connections of the different electrodes (EDA & ECG) and reference points as well as to validate the quality of the signals.



*Figure 2: BITalino biomedical data acquisition dev board*

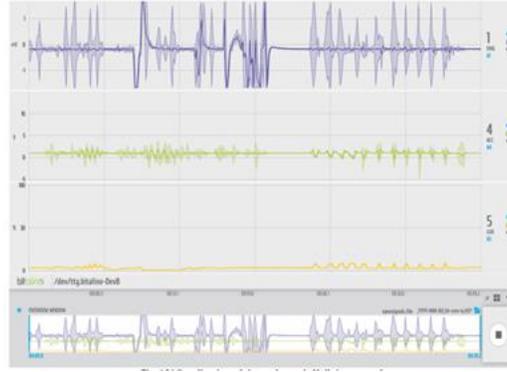


Figure 3: OpenSignals software for data acquisition, visualization and processing of physiological signals

A Bluetooth connection was setup between Bitalino and Opensignals. An output marker in Opensignals in a form of a square wave was used to indicate the start and the end of each measurement. Two different laptops were used, one for EEG and the other for physiological information, the start of the measurement of physiological information were manually synchronized with the measurements of EEG signals by pressing the start input markers at the same time.

Furthermore, a python code was developed to extract the measurement sections while relying on the input. The codes also convert the data from OpenSignals to significant physiological units using the following transfer functions.

For EDA:

$$R_{MOhm} = 1 - \frac{EDA_B}{2^n} \quad EDA_{\mu S} = \frac{1}{R_{MOhm}} \quad (1)$$

Where  $EDA_{\mu S}$  is the EDA value in micro Siemens

And  $EDA_B$  is the EDA value obtained from Bitalino

$R_{MOhm}$  is the sensor resistance in megaOhms

n number of bits (bit)

For ECG:

$$ECG_V = \frac{\frac{ECG_B \times V_{CC}}{2^n} - \frac{V_{CC}}{2^n}}{G_{ECG}} \quad (2)$$

$ECG_V$  ECG value in Volts

$ECG_B$  ECG value obtained from Bitalino

$V_{CC}$  Operating voltage (V)

n number of bits (bit)

$G_{ECG}$  ECG sensor Gain

### 3.2.5. Ledalab

Ledalab is Matlab based software for the analysis of EDA (Benedek & Kaernbach 2010a) (Benedek & Kaernbach 2010b). It performs event-related analysis relative to events/markers and return various parameters of tonic and phasic activity. Using the different signal measurement markers, Ledalab was used to preprocess and smooth the measured EDA data, and to analyze the signal using its decomposition analysis options. Specifically, the Continuous Decomposition Analysis is used, as in this thesis, the physiological information is measured in a short duration, thus this decomposition analysis takes into account all the data instead of just doing peak analysis. Furthermore, it is robust to movement artifacts.

## 4. Method

### 4.1. Data collection and experiment procedure

Before the experiment started, participants were informed about the details of the experiment. The subjects were informed to sit in a comfortable position, to keep their eyes closed and to not move during the experiment. The Emotiv EPOC was placed on the subject's scalp and guaranteed that all the electrodes are connected and working properly. Additionally, skin conductance electrodes were placed on the subject's palms. Two additional electrodes were placed on the subject's wrist to measure heart rate.

EEG data in this thesis was collected from 3 healthy male participants. Subjects listened to 40 tracks from the dataset (8 from each discrete emotion) with a duration of 12 seconds each. The tracks were randomized, however 2 tracks from the same emotion category were placed successively. After each 2 tracks, the subject reported their induced emotions in terms of pleasantness, energy, tension, anger, fear, happiness, tenderness and sadness using a liKert questionnaire. Due to the fact that in this study we adopted a discrete model for emotions, the rating of the targeted emotion was only investigated as to validate that the induced emotion is similar to the label of the track in the original dataset.

### 4.2. EEG Feature extraction

The EEG data was acquired with an internal sampling rate of 128 Hz. To remove linear trends and power line noise, all EEG signals were filtered using a high pass filter at 1 Hz and low passed at 50 Hz using an FIR filter as with the recommendation of EEGLAB (Delorme & Makeig 2004). The EEG signals were band-pass filtered using a 4<sup>th</sup> order butterworth filter to extract 5 band waves: delta ( $\delta$ : 1–4 Hz), theta ( $\theta$ : 5–8 Hz), alpha ( $\alpha$ : 8–12 Hz), beta ( $\beta$ : 12–32 Hz), and gamma ( $\gamma$ : 31–39 Hz). The feature includes log power of each of the 14 electrodes in 5 band waves. The log power is expressed by the following equation:

$$LPf = 1 + \log\left(\frac{1}{N} \sum_{n=1}^N (x_{nf})^2\right) \quad (3)$$

where  $x_{nf}$  is the magnitude of the  $f$ th frequency band of the  $n$ th sample,  $N$  is the number of samples.

Additionally, 2 log power asymmetry features in 5 band waves were extracted, which are defined by the difference between F3 and F4, AF3 and AF4.

Moreover, 2 additional features were extracted which we represent by Arousal and Valence.

$$\text{Arousal} = \frac{\beta F3 + \beta F4 + \beta AF3 + \beta AF4}{\alpha F3 + \alpha F4 + \alpha AF3 + \alpha AF4} \quad (4)$$

$$\text{Valence} = \frac{\alpha F4}{\beta F4} - \frac{\alpha F3}{\beta F3} \quad (5)$$

where  $\alpha$  and  $\beta$  are the logarithmic power of a certain electrode in the alpha and beta bands respectively.

*Table 3: EEG extracted features*

Feature type	Number of features
<b>Log power</b>	70
<b>Log power Asymmetry</b>	10
<b>Arousal</b>	1
<b>Valence</b>	1
<b>Total</b>	82

The EEG features were extracted using a 1 second window size and 0.5 second hop size, thus the number of samples from each subject was 920 points (23 frames for each 12 seconds track x 40 tracks)

### 4.3. Audio features extraction

This study employed ESSENTIA (Essentia & Gaia) to extract audio features that represent various perceptual dimensions in music, including pitch, loudness, dissonance, rhythm and timbre. Additionally, all extracted features were picked to be significant in terms of representing the acoustic properties of an audio segment in our frame based approach. A total of 30 features were extracted.

Loudness measures the perceived intensity of sounds. Loudness is regarded as being relevant to express emotions (Juslin & Laukka, 2004).

Consonance and dissonance are seen as being relevant in emotion perception (Koelsch et al. 2006) Dissonance measures the degree of roughness of the acoustic spectrum (Sethares 2005). Consonant sounds have evenly spaced peaks compared to dissonant sounds.

(Juslin & Laukka, 2004) reported that there is also a relationship between timbre and emotions. Therefore, we extracted 13 MFCC coefficients. MFCCs characterizes the spectral shape of an acoustic spectrum. MFCCs are computed by performing a discrete cosine transform of log-power spectra expressed on a non-linear perceptual-related Mel-frequency scale.

Pitch salience measure the *pitchness* of a sound. The pitch salience is computed by considering the ratio of the highest auto correlation value of the spectrum to the non-shifted autocorrelation value(Ricard 2004). Pitch salience represents a measure of tone sensation. This ratio has a value close to 1 for harmonic sounds, and close to 0 for non-harmonic sounds.

Three spectral features were extracted: Spectral centroid, spectral flux and spectral roll-off. The spectral centroid defines the center of gravity of a magnitude spectrum and it identifies where the concentration of energy is most prominent. Spectral centroid provides a measure of a brightness of a sound. Spectral flux provides a measure of the rate of change of the spectrum of a sound. Spectral roll-off measures the tonality of a signal, as it is defined as the frequency below which 95 % of the power spectrum is located.

The modes of the chords and their strength features were also included, as musical modes are reported as distinctive among different emotions (major for happy and tender, minor for sad) (Gabrielsson & Lindström 2010).

There are numerous features of music that have been reported to be relevant to express discrete emotions (Juslin & Laukka, 2004). Rhythm is one of the features that have distinct characteristic among different discrete emotions, for example, fast tempo in happy songs compared with slow tempo in sad and tender songs.

For our frame based approach, onset rate was extracted from the audio tracks as a feature to express the rhythmic composition of an excerpt. The onset rate is an estimation of the number of events occurring in one second.

All audio features were extracted with a 1 second window size, and a 0.5 second hop size, similarly to the features extracted from EEG. The features and their statistics are summarized in the table below.

*Table 4: Features extracted from audio*

<b>Feature type</b>	<b>Number of features</b>
<b>Dissonance</b>	2 (mean and variance)
<b>MFCC</b>	13 coefficients
<b>Pitch salience</b>	2 (mean and variance)
<b>Spectral centroid</b>	2 (mean and variance)
<b>Spectral flux</b>	2 (mean and variance)
<b>Spectral roll-off</b>	2 (mean and variance)
<b>Zero crossing rate</b>	2 (mean and variance)
<b>Onset rate</b>	1
<b>Chords scale</b>	1
<b>Chords strength</b>	2 (mean and variance)
<b><i>Total</i></b>	<b><i>30</i></b>

#### 4.4. Fusion of EEG and audio features

In multidisciplinary approaches, the fusion of two single modalities is usually achieved by either a decision level fusion or through a feature level fusion. A decision level fusion scheme processes the single modalities independently and the output of the classifiers are combined to yield the final result depending on a weight factor that determines the degree of contribution of each modality. Feature level fusion concatenates the features of both modalities into one composite feature vector, and the vector is then fed to the classifier. In this study, as the objective is to assess the degree of contribution of audio features to the classification accuracy of the multimodality, we adopted a feature level fusion. The EEG and audio features are combined to form a vector of 112 features (82 EEG features and 30 audio features). Both features types were extracted using a 1 second window size and a 0.5 second hop size. For each subject the multimodal vector is composed by 920 samples (23 frames for each 12 seconds track x 40 tracks). All the features of the single modalities were independently normalized between 0 and 1, thus making the features equally weighted.

#### 4.5. Skin conductance features extraction

Skin conductance is characterized by 2 main components: The tonic and phasic components. The tonic component (i.e. Skin conductance level, SCL) is the slowly varying baseline level of SC and it is related to a person's overall state of arousal. The phasic component refers to the faster changing elements of the signal. It might arise within a predefined response window (1-5 s after stimulus) (Benedek & Kaernbach 2010a).

The figure below shows the two different drivers computed from the raw skin conductance signal.

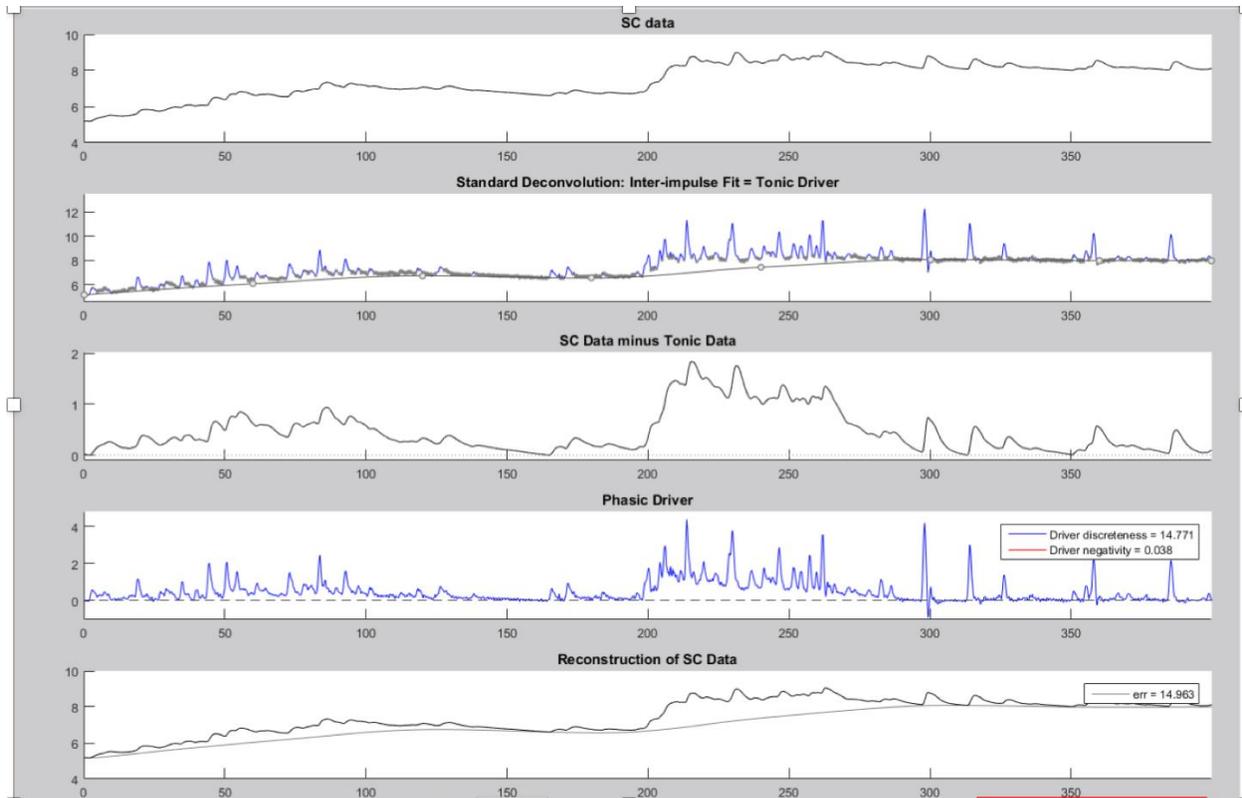


Figure 4: The main components of skin conductance computed from the raw signal using decomposition analysis

Ledalab (Benedek & Kaernbach 2010a) (Benedek & Kaernbach 2010b) is used to perform the decomposition analysis in addition to extracting different features depicted in the table below.

The batch mode in Ledalab was used to extract the various features from the raw input SC signal. The signals are preprocessed using a 1<sup>st</sup> order low-pass Butterworth filter with a 5Hz cutoff frequency. Down sampling is also applied to speed up the computation. Furthermore, smoothing using a hanning window with a 4 samples width is applied to the raw signal in addition to optimizing the parameters of the decomposition.

To extract the skin conductance features, a 4 second window size with no overlap is used. A frame based approach is adopted. Hence, the number of samples for each 12 second track will be 3 samples. The total number of samples for a particular subject is 120 samples (3 samples for each track times a total of 40 tracks)

Table 5: Skin conductance extracted features and their description

Feature	Description
<b>Continuous Decomposition Analysis (CDA)</b>	
<b>CDA.nSCR</b>	Number of significant (= above-threshold) SCRs within response window (wrw)
<b>CDA.AmpSum</b>	Sum of SCR-amplitudes of significant SCRs wrw (reconvolved from corresponding phasic driver-peaks) [ $\mu\text{S}$ ]
<b>CDA.SCR</b>	Average phasic driver wrw. This score represents phasic activity wrw most accurately, but does not fall back on classic SCR amplitudes [ $\mu\text{S}$ ]
<b>CDA.JSCR</b>	Area (i.e. time integral) of phasic driver wrw. It equals SCR multiplied by size of response window [ $\mu\text{S}*\text{s}$ ]
<b>CDA.PhasicMax</b>	Maximum value of phasic activity wrw [ $\mu\text{S}$ ]
<b>CDA.Tonic</b>	Mean tonic activity wrw (of decomposed tonic component)
<b>Standard trough-to-peak (TTP) or min-max analysis</b>	
<b>TTP.nSCR</b>	Number of significant (= above-threshold) SCRs within response window (wrw)
<b>TTP.AmpSum</b>	Sum of SCR-amplitudes of significant SCRs wrw [ $\mu\text{S}$ ]
<b>Global Measures</b>	
<b>Global.Mean</b>	Mean SC value within response window (wrw)
<b>Global.MaxDeflection</b>	Maximum positive deflection wrw

#### 4.6. Electrocardiogram (ECG) features

To extract features related to ECG and heart rate, Kubios software was adopted (Tarvainen et al. 2014). Due to the software requirements of using a minimum 30 seconds of heart signals, the ECG in the experiment was measured during two consecutive tracks from the same emotion class to achieve the 30 seconds window. The following figure shows the output analysis of one 30 seconds sample from the raw ECG signals. This snapshot shows the different features computed, time domain features such mean heart rate, standard deviation of heart rate peaks, and other frequency domain features related to the spectral power in different frequency bands.

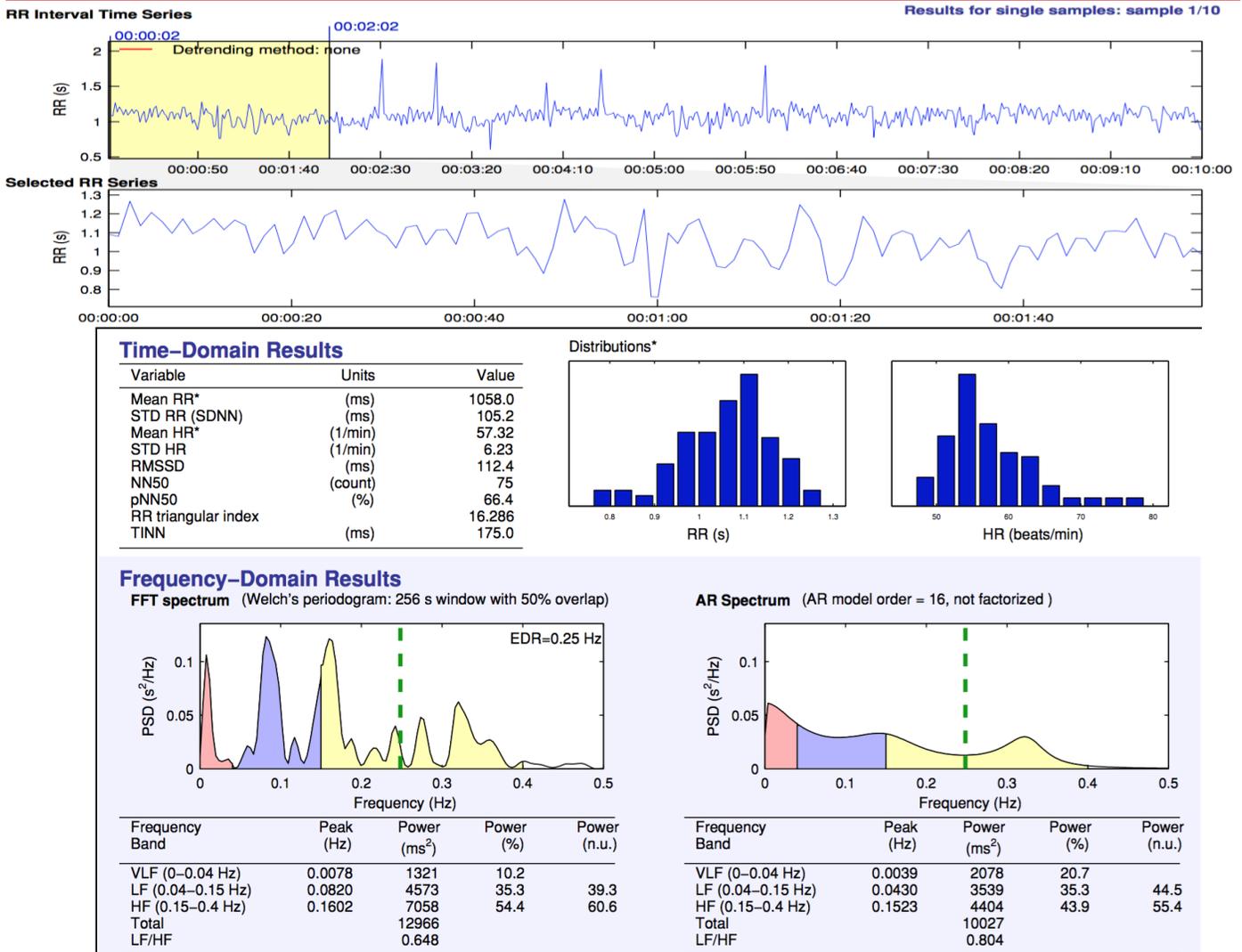


Figure 5: Output analysis of the Kubios software of a sample, and all the features extracted.

The final ECG vector size is 20 samples (There is 40 tracks, combining each two tracks from the same emotion class to form a 30 seconds track)

#### 4.7. Fusion of EEG and skin conductance

A feature level fusion was adopted for a multimodal approach combining EEG and skin conductance features. The EEG and skin conductance features were combined to form a vector of 92 features (82 EEG features and 10 SC features) and 120 samples (3 samples for each track times a total of 40 tracks).

The window size is 4 second with no overlap. All the features of the single modalities were independently normalized between 0 and 1, thus making the features equally weighted.

## 4.8. Feature Classification

### 4.8.1. EEG classification

Two algorithms were evaluated for the EEG classification. K-nearest neighbors algorithm with different k values and two distance metrics are evaluated, and Support Vector Machine (SVM). The K-NN algorithm, given a query point, finds the k closest to the query point (Cover & Hart 1967). Support vector machine (SVM) is a well-known and popular machine learning algorithm and it is widely used in EEG emotion classification. SVM projects input data onto a higher dimensional feature space via a kernel function. Moreover, SVM constructs a separating hyperplane between two classes, the positive and negative samples, in such a way that the distance from each of the two classes to the hyperplane is maximized. LIBSVM library in Weka was used (Chang & Lin 2013). A grid-search approach was applied to individual datasets to decide an optimal parameter pair ( $\gamma$ , C) of RBF kernel. We adopted the range ( $\gamma$ :  $2^{-3} \sim 2^3$ , C:  $2^{-3} \sim 2^3$ ).

Additionally, in order to compare the classification accuracy using all EEG features with an automatic feature selection algorithm, we adopted Wrapper feature selection using a KNN searching algorithm and we compared the results with the classification using all features.

All classification accuracies in this thesis were computed using 10 runs of 10 cross-validation, and a *t*-test was performed with a significance value of 0.05.

### 4.8.2. Multimodal classification (EEG and Audio)

For this multimodal classification, we adopted the SVM algorithm with the parameters of ( $\gamma$ , C) pair computed using grid search in the previous EEG classification task. The SVM was adopted as it is widely used in EEG and multimodal emotion classification, and this will enable us to compare our results with other research using this algorithm for emotion and perception related tasks. Furthermore, we investigated the classification accuracies with varying number of features

as to assess the addition of audio features in the classification problem. Additionally, we studied the effect of varying the training size on the classification accuracy.

#### 4.8.3. Multimodal classification (EEG and skin conductance)

Due to the large number of features compared with the size of the samples, 92 features (82 EEG features and 10 SC features) and 120 samples, feature selection is adopted as to reduce the dimensionality of the classification task and reduce possible overfitting.

The Relief feature selection method was adopted. This method evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class (Kira & Rendell 1992)(Kononenko 1994).

Support Vector Machine (SVM) was adopted in this classification task.

## 5. Results

### 5.1. EEG mood classification

The results for the EEG classification show that the best results were obtained using the k-NN algorithm with k=1 with Manhattan distance as a distance measurement metric.

The table below shows the EEG classification accuracies for the three subject using 10 runs of 10 cross validation and a *t*-test with a 0.05 significance value.

Table 6: EEG mood classification results.(v) Indicates that the accuracy using k-NN significantly outperformed SVM

	KNN	SVM
<b>All features</b>		
<b>EEG Subject1</b>	87.03% (v)	76.70%
<b>EEG Subject2</b>	89.75% (v)	81.35%
<b>EEG Subject3</b>	83.84% (v)	77.23%
<b>With feature selection (Wrapper)</b>		
<b>EEG Subject1 (41 features)</b>	87.15% (v)	79.82%
<b>EEG Subject2 (35 features)</b>	88.02% (v)	83.11%
<b>EEG Subject3 (46 features)</b>	89.24% (v)	83.61%

The graphs below compare the classification accuracies of the k-NN algorithm for the three subjects using different k values and two distance metric: Euclidean and Manhattan distance.



Figure 6: Comparison of k-NN classification accuracy for different k values and two distance metrics for three participants.

It is important to evaluate the effect of feature selection in selecting the most informative features in the classification task. Thus we compared the EEG mood classification using the wrapper feature selection for different sizes of the dataset.

*Table 7: Comparison of EEG mood classification using all features with an approach using Wrapper feature selection. (v) indicates that the accuracy using wrapper feature selection significantly outperforms the classification using all features*

	All features (SVM)	Wrapper feature selection (SVM)
<b>EEG Subject1</b>	76.70%	79.82% (v) (41 features)
<b>EEG Subject2</b>	81.35%	83.11% (v) (35 features)
<b>EEG Subject3</b>	77.23%	83.61% (v) (46 features)

The below graph compares both the classification using all features, with the classification using wrapper as feature selection. Different percentages of the dataset are investigated. We can see that using the wrapper feature selection, the accuracy is always higher than the classification using all the features for all dataset sizes.

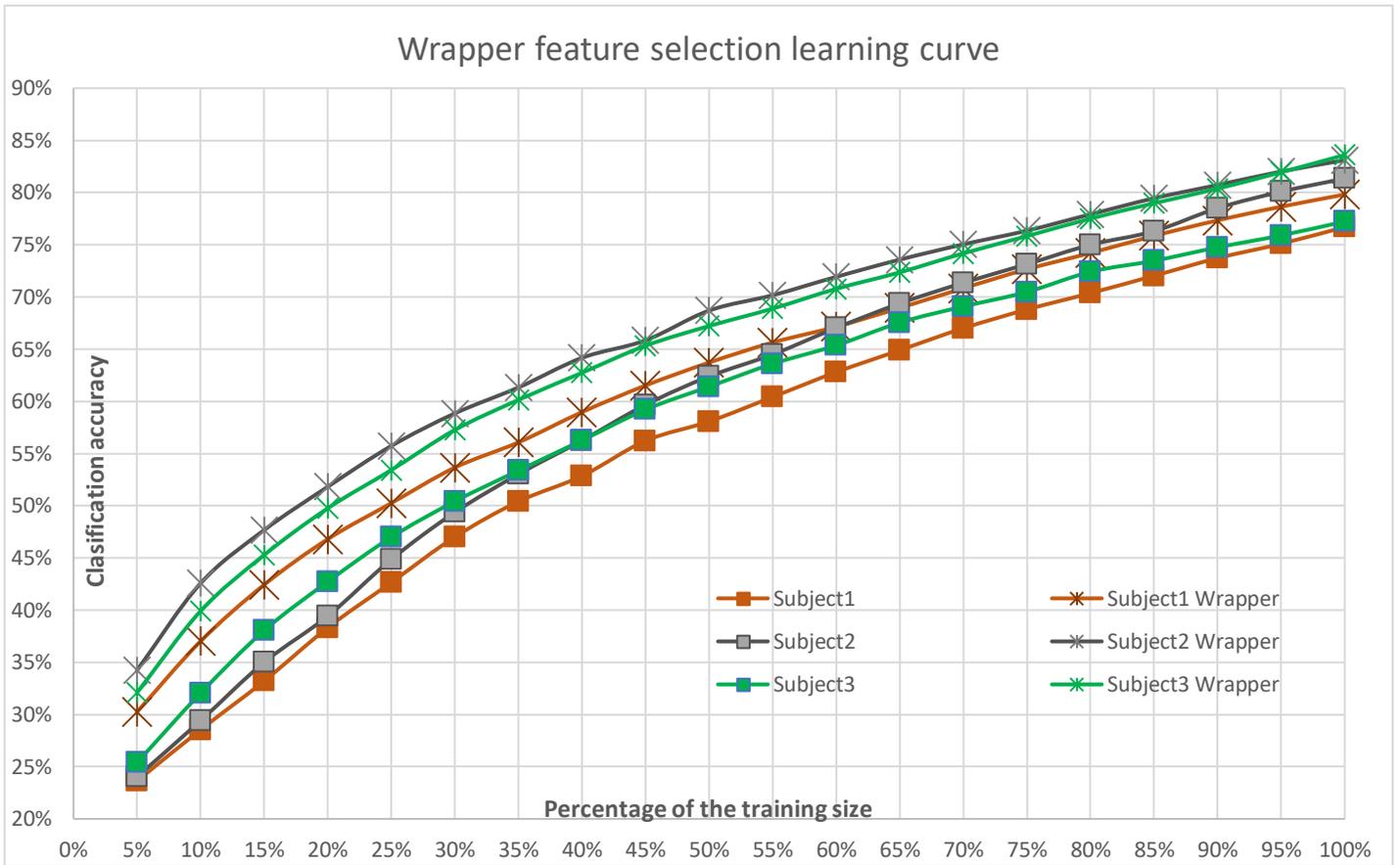


Figure 7: Comparison of EEG classification accuracy with and without feature selection for different training/testing size.

## 5.2. Multimodal classification (EEG and Audio features)

### 5.2.1. Multimodal classification accuracy

In order to assess the degree of contribution of audio features to the classification accuracy of the multimodality, we adopted a feature level fusion.

The EEG and audio features are combined to form a vector of 112 features (82 EEG features and 30 audio features). For each subject the multimodal vector is composed by 920 samples (23 frames for each 12 seconds track x 40 tracks). We adopted the SVM algorithm with the parameters of ( $\gamma$ , C) pair computed using grid search in the previous EEG classification task. The table below shows the classification accuracies of the three subjects for EEG alone and the multimodal approach. The accuracies were computed using 10 runs of 10-fold cross validation in addition to a  $t$ -test with a

0.05 significance value. We can clearly see that the multimodal approach outperformed the stand-alone EEG classification for all three subjects.

*Table 8: Multimodal (EEG and Audio) classification results for three participants (v) Indicates that the accuracy using the multimodal approach significantly outperforms the classification using EEG as a stand-alone method.*

<b>SVM</b>	
<b>All features</b>	
<b>EEG Subject1</b>	76.70%
<b>EEG Subject2</b>	81.35%
<b>EEG Subject3</b>	77.23%
<b>EEG and Audio features</b>	
<b>Multimodal Subject1</b>	85.59% (v)
<b>Multimodal Subject2</b>	89.11% (v)
<b>Multimodal Subject3</b>	88.21% (v)

Furthermore, to ascertain that the inclusion of audio features does not produce an arbitrary increase in the classification accuracies resulting from adding features to the classification approach, we investigated the classification accuracies while varying the total number of EEG features. The graph below displays the behavior of the classification accuracies for the three subjects while increasing the number of features. We must mention that the first 82 features added to this assessment are only EEG features. From the graph we can notice that around 50 features, the classification accuracies for the three subjects becomes more or less stable with no big increase in the accuracy. However, when we start to add the audio features (The audio features starts at the index value of 83) the classification accuracy starts to significantly increase. This investigation validates that the inclusion of audio features and its effect on increasing the classification accuracy is not an arbitrary effect of just adding features to the classification task, as the graph shows that at some point, the trend line of the accuracy stabilizes.

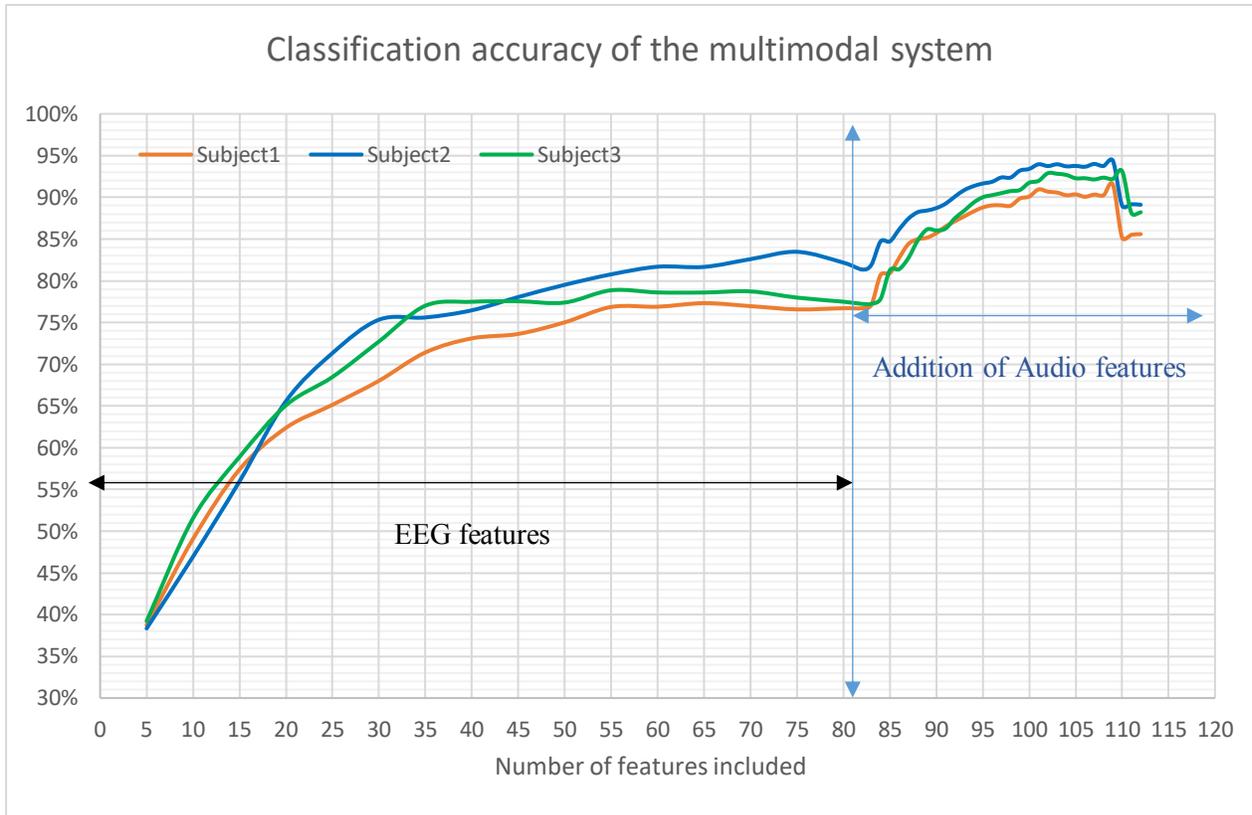


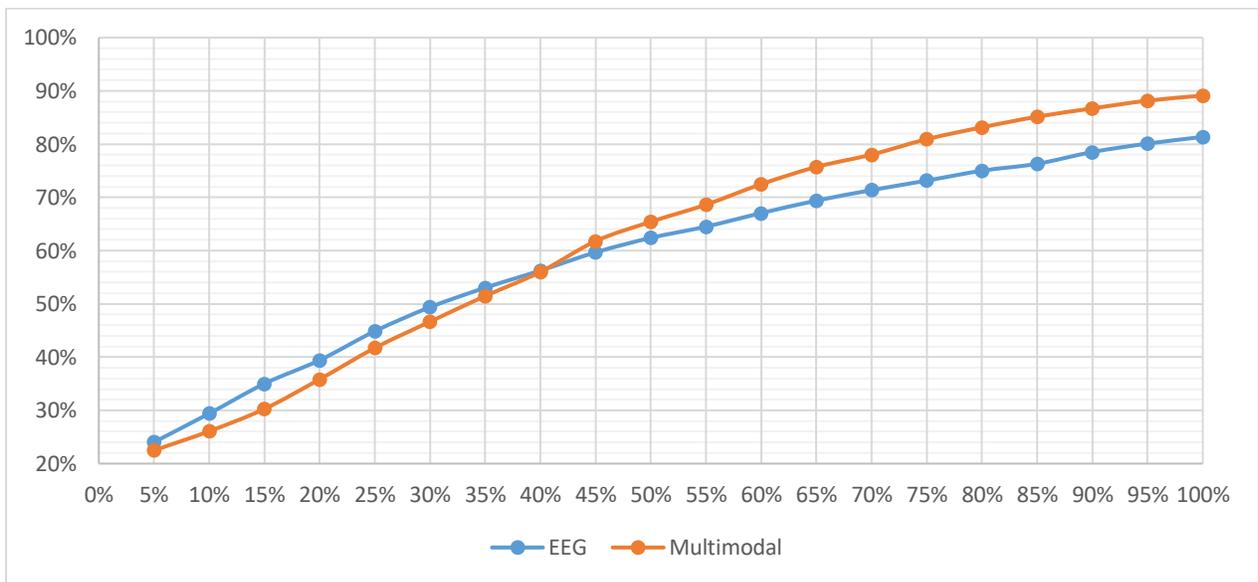
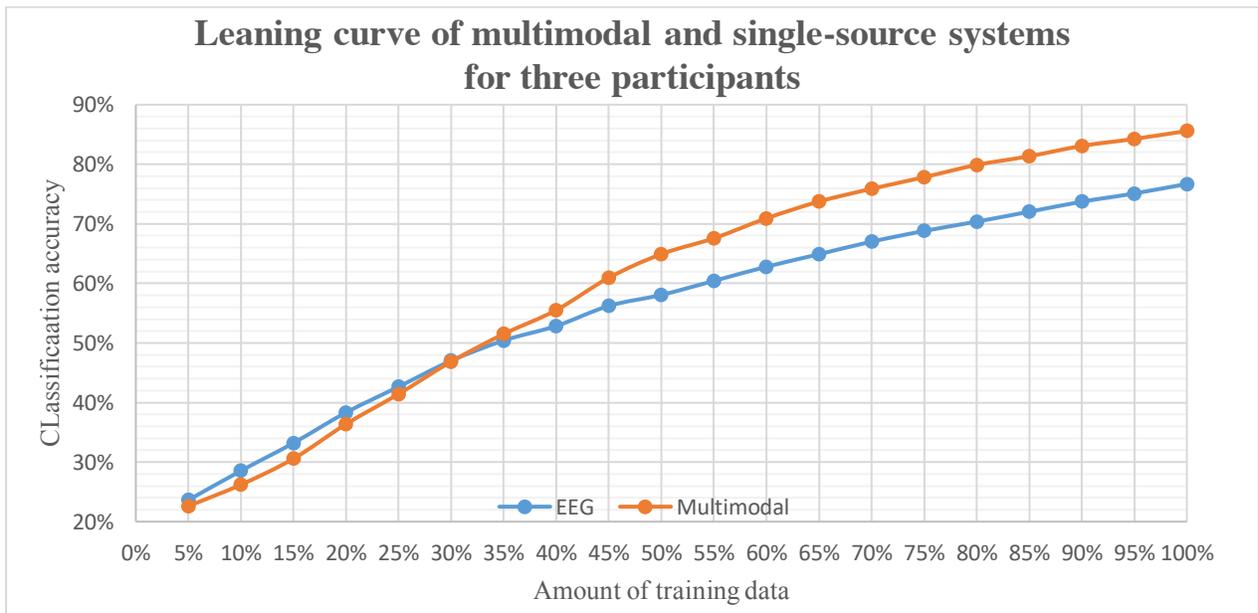
Figure 8: Multimodal classification accuracy, before and after the inclusion of Audio features.

### 5.2.2. Learning curves

A learning curve describes the relationship between the classification performance and the size of the dataset used in the training examples. Usually the performance of a classifier increases with increasing the number of training examples, and at the point where the performance stops increasing indicates the minimum number of training examples needed to achieve the optimal accuracy and performance. Additionally, learning curves are important in assessing the performance of a classifier in a multimodal system, in which it evaluates the effectiveness of the classifier with comparison to a mono-modality system and can reveal whether the hybrid system can reduce the number of training examples to achieve a comparable or even better performance than stand-alone systems.

The following graphs compares the performance of the multimodal system combining EEG and Audio features with the EEG stand-alone system for the three subjects with a varying number of training examples. The evaluation was computed using 10 runs of 10 cross validation.

For a small number of training examples, between 25% and 40% of the original training dataset, the performance of the single source systems outperformed the multimodal system. This is might be due that Support Vector Machine (SVM) need a suitable size of training examples to output a good performance. However, with greater number of training examples, the multimodal system clearly outperforms the mono-modal system. We can notice that the multimodal system is able to reduce the number of training examples from 25% to 35% compared with the stand-alone EEG system to achieve a comparable classification accuracy.



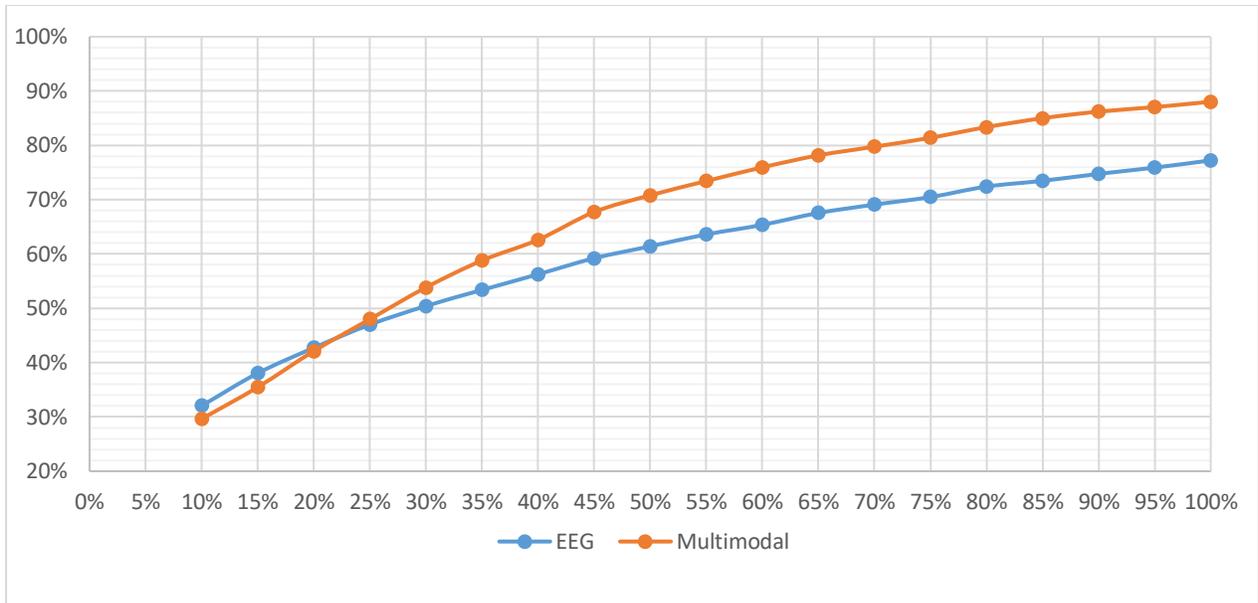


Figure 7: Learning curve of multimodal and single-source systems for three participants.

### 5.2.3. Contribution of individual Audio features

To assess the contribution of audio features in the classification performance of the multimodal system, each audio feature was individually added to all of the EEG features and the classification accuracy is computed. The evaluation was performed using 10 runs of 10-fold cross validation and SVM as the machine learning algorithm with the with the parameters of  $(\gamma, C)$  pair computed using grid search in the previous EEG classification task. The below chart illustrates the contribution of each audio feature in terms of how much the overall classification accuracy increases or decreases after the inclusion of the correspondent audio feature.

For the three subjects, spectral dissonance shows a noticeable increase in the classification accuracy by 2.82%, 3.72% and 3.78% for the three subjects. 0<sup>th</sup> and 1<sup>th</sup> MFCC coefficients effectively increased the accuracy of the system. As we go higher in the MFCC coefficient, the percentage of contribution of the coefficients decreases, except for 5<sup>th</sup> and 6<sup>th</sup> MFCC coefficients which individually and noticeably increased the classification accuracy especially for the third subject by more than 4%.

Concerning pitch and tonality, the mean of pitch salience showed a big contribution to the overall accuracy above 3% for each of the three subjects.

The spectral audio features also showed their influence in increasing the classification accuracy. Specifically, the spectral centroid, spectral flux and spectral roll-off individually improved the accuracy between 2% and 4%.

Concerning rhythmic features, zero crossing rate and onset rate were proven to be useful in this classification task.

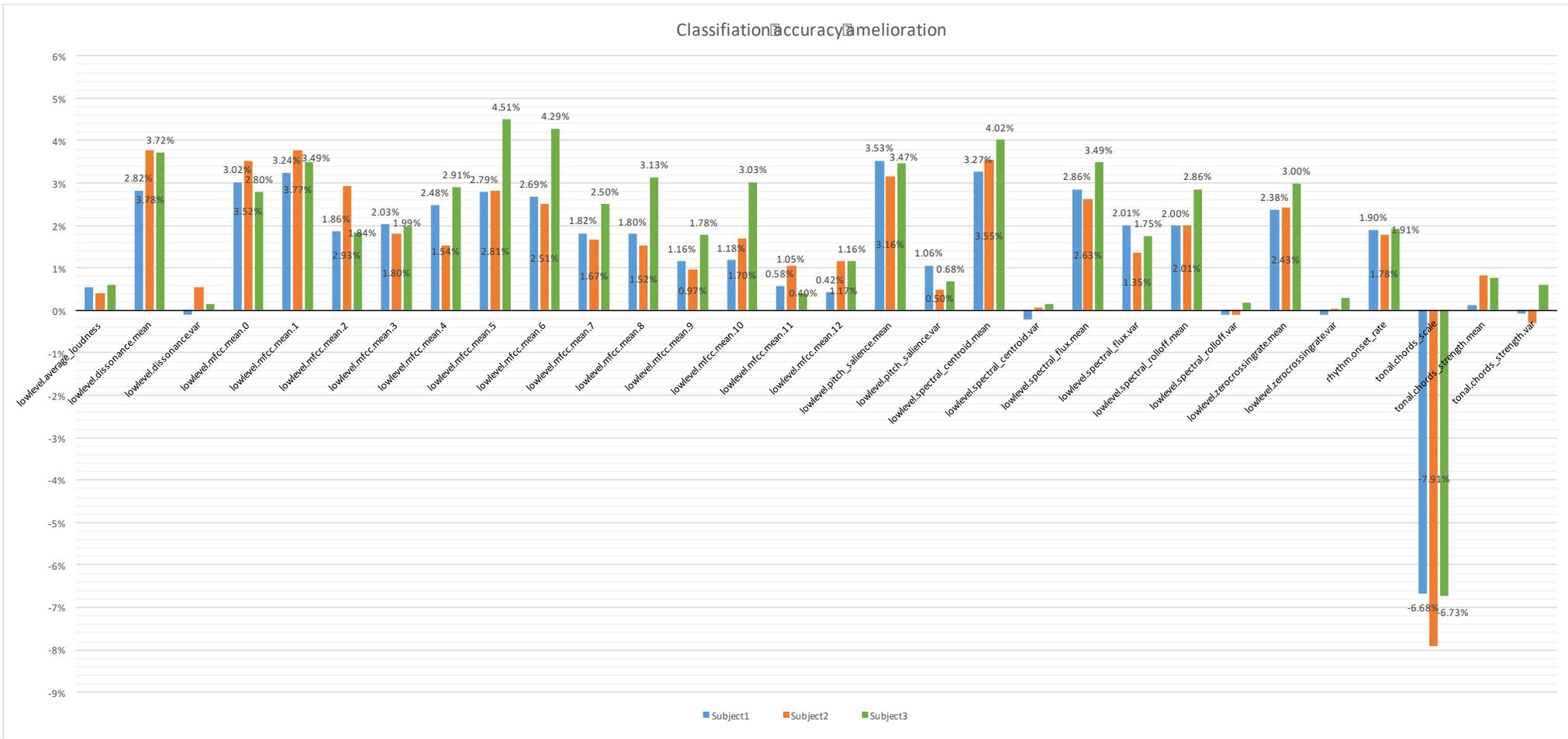


Figure 10: EEG Classification accuracy amelioration after the inclusion of individual audio features

#### 5.2.4. Comparison

Comparing with (Laurier et al. 2009), in which they explored the relationship between audio features and emotions in music. The work is based on 110 excerpts from film soundtracks created in a study by (Eerola & Vuoskoski 2011) evaluated by 116 listeners. The data is annotated with 5 basic emotions (fear, anger, happiness, sadness, tenderness) on a 7 points scale. They extracted audio features and investigated the correlation between the features and the 5 discrete emotions.

A positive correlation was found between dissonance and fear and anger in addition to a negative correlation with sad and tenderness.

Onset rate was found correlated with the happy category whereas loudness was found correlated with happy and anger.

Both of the dissonance and the onset rate were also proven in our study to successfully increase the classification accuracy after their addition to the EEG features.

In (Daly et al. 2015), they found that features related to the variance of Mel cepstral coefficients in high frequency bands was related to valence. Low frequency Mel cepstral coefficient was found related to Energy-arousal and Tension-arousal.

In our study, low coefficient MFCCs (0<sup>th</sup> and 1<sup>th</sup>), were found to be efficient in increasing the classification accuracy, in addition to MFCC 5<sup>th</sup> and 6<sup>th</sup>.

(Trochidis et al. 2011) showed that low level spectral and temporal features such as flatness and chroma are efficient in modeling the arousal dimension, whereas high-level features such articulation, pulse clarity, mode and brightness succeed to measure the valence dimension. In our study, spectral centroid which can represent the brightness of a sound was found to increase the classification accuracy.

Finally, (Vempala & Russo 2012) adopted neural networks for predicting arousal/valence ratings from participants listening to 12 classical music excerpts from 12 classical composers. They found strong positive correlations between arousal and five audio features: pulse clarity, zero crossing rate, spectral centroid, spectral roll-off and brightness. Low energy and mode we found positively correlated with valence.

In our study, zero crossing rate, spectral centroid, spectral roll-off and MFCC 0<sup>th</sup> coefficient were also found to noticeably increase the classification accuracy when added to the EEG features.

The results are summarized in the table below.

*Table 9: Table comparing the results of our study with previous research dealing with correlation between audio features and emotion rating from participants*

Author	Type of study	Dataset	Features
<b>Laurier et al. 2009</b>	Correlation between audio features and emotion ratings	Same dataset	Dissonance Onset rate Loudness
<b>Daly et al. 2015</b>	Fusion of audio and EEG	Same dataset	Variance of MFCCS
<b>Trochidis et al. 2011</b>	Correlation between audio features and emotion ratings	Contemporary Western music	Flatness, Chroma spectral centroid , mode and brightness
<b>Vempala &amp; Russo 2012</b>	Correlation between audio features and emotion ratings	Classical music excerpts	pulse clarity, zero crossing rate, spectral centroid, spectral roll- off and brightness

### 5.3. Multimodal classification (EEG and skin conductance)

The Relief feature selection method was adopted in the classification task. This method evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class (Kira & Rendell 1992)(Kononenko 1994) and then ranks the features. We decided to include 12 features from the multimodal system. In both experiments, CDA tonic and Global mean skin conductance features were selected by the feature selection algorithm with the highest ranking. The multimodal system consists of 10 EEG features and 2 skin conductance features. Support Vector Machine (SVM) was adopted as a classification algorithm. Given that the window size used in this classification task is 4 seconds

with no overlap, hence there is 120 samples (3 samples for each track times a total of 40 tracks) to be included.

The results are shown in the table below. Two skin conductance features, CDA tonic and Global mean were noticeably able to increase the classification performance by 26.7% and 18.3% for the two subjects compared with the stand-alone EEG system.

*Table 10: Multimodal classification results (EEG and Skin conductance)*

<b>Multimodal classification</b>	
<b>EEG features</b>	
<b>EEG Subject1</b>	35.8 %
<b>EEG Subject2</b>	43.3 %
<b>EEG and Skin conductance features</b>	
<b>Multimodal Subject1</b>	62.5 %
<b>Multimodal Subject2</b>	61.6%

#### 5.4. ECG measurement

For Heart Rate Variability (HRV), and due to small number of samples that is considered insufficient for performing a classification task, the correlation of the features with the emotions is calculated as to provide an initial assessment of using HRV features computed in a short duration. The correlation was computed using WEKA.

From two subjects, the very low frequency peak of the HRV signal showed a good correlation with the emotion classes (0.24 and 0.3). Additionally, time domain features, such as the mean heart rate and the standard deviation of the heart rate peaks are also correlated with the emotion classes giving that a short duration measurement is used. Finally, low frequency to high frequency ration was shown to be correlated to the emotion classes in both subjects.

Table 11: Correlation of ECG features with the emotion classes.

<b>ECG correlation</b>		
<b>Frequency domain features</b>		
<b>Subject1</b>	VLF_peak	0.2412
<b>Subject2</b>	VLF_peak	0.3
<b>Time domain features</b>		
<b>Subject1</b>	mean_HRV	0.2057
<b>Subject2</b>	std_RR	0.338
<b>Ratio between LF and HF band powers</b>		
<b>Subject1</b>	LF_HF_power	0.225
<b>Subject2</b>	LF_HF_power_welsh	0.2294

## 6. Discussion

Discussion is divided into three parts. In the first part, the results of the EEG mood classification will be discussed. In the second part, the multimodal classification will be assessed. And finally, in the last part, the mood classification using physiological signals will be summarized.

### 6.1. EEG mood classification

The results for emotion recognition from the stand-alone EEG approach were found quite satisfying. K-NN classification was investigated for different k values and different distance metrics. K-NN algorithm with k=1 and Manhattan distance performed best, with accuracies of 87.03%, 89.75% and 83.84% for three male participants.

However, SVM is widely used in research dealing with EEG signals, so we adopted this algorithm as to be able to compare with previous research. SVM accuracies for EEG emotion recognition were 76.70%, 81.35%, and 77.23% for the three participants. We also found that when using feature selection, specifically the wrapper feature selection, we were able to obtain higher accuracies using fewer number of features. The accuracies using features selection increased to 79.82%, 83.11% and 83.61% with about half the number of features compared to the previous approach.

Moreover, learning curves of the EEG classification approach were investigated for the SVM algorithm. For different training set size, the wrapper feature selection always outperformed the all-features approach for all the subjects. We can conclude that with fewer features, we were able to obtain higher accuracies given that the nature of EEG signals are noisy and dependent on each subject, thus a refined set of informative EEG features from each participants would improve the classification accuracy over a system using all features.

Investigating the performance of the algorithms, knn showed better classification accuracies, especially with k=1, and is mainly due to that some emotion categories such fear and angry share several similarities, thus the confusion between samples from these two categories were minimized when using 1 nearest neighbor instead of the hyperplane of the SVM.

Finally, these results show that it is efficient to use a categorical model of emotion for emotion recognition from EEG, as the classification accuracies are comparable with other research using a dimensional model of emotions.

## 6.2. Multimodal classification

A feature level fusion was adopted to combine features from both EEG and Audio. The results show that the multimodal system outperformed the EEG mono modal system. The multimodal system achieved classification accuracies of 85.59%, 89.11% and 88.21% for the three participants. Furthermore, to make sure that the increase in classification accuracies is not random and only related to addition of features, we investigated if the increase in the accuracies is related to the type of features added. The assessment concludes that the addition of audio features noticeably increased the classification accuracies, in contrast to a saturated accuracy after the addition of different EEG features. These results concludes that there is some relevant and important acoustic information in the audio features which could improve the performance of the system.

Furthermore, we investigated the performance of the multimodal system by plotting the learning curves of the system. The curves showed that, for a suitable amount of training and test sizes, the multimodal system outperformed the mono modal system for different set sizes. Moreover, the multimodal system was able to reduce the number of training examples from 25% to 35% compared with the stand-alone EEG system to achieve a comparable classification accuracy.

However, for a small number of training and test sizes, the stand alone EEG system outperformed the multimodal system, the reason is mainly due to that the SVM algorithm, and for two different feature types, requires a good amount of training set to be able to decide on a good hyperplane separating the emotion classes.

Additionally, we evaluated the contribution of each audio features in the classification performance of the multimodal system, each audio feature was individually added to all EEG features and the classification accuracy is computed. Several audio features noticeably increased the classification accuracy after their addition. Comparing with previous research dealing with correlation between audio features and emotion rating using the same or different datasets, we

found features in common that were also found to be informative, such as, dissonance, onset rate, MFCCS, spectral centroid, zero crossing rate and spectral roll off. Detailed classification accuracy amelioration of each audio features are found in figure 10. Table 9 summarizes the information of previous research we used for comparison.

### 6.3. Multimodal Classification using physiological signals

When combining EEG and skin conductance features, and when using the Relief feature selection method, using only 12 features (10 from EEG and 2 from skin conductance), the multimodal system achieved 62.5% and 61.6% classification accuracies for two subjects, Moreover, two skin conductance features were able to increase the classification accuracy by 26.7% and 18.3% for two subjects compared with the stand-alone EEG system. These results shows that the activation of the autonomic nervous system changes when emotions are elicited and that physiological signals provides an opportunity to recognize affect changes that cannot be easily perceived visually and from EEG signals.

Furthermore, from Heart Rate Variability (HRV) measured in a very short duration, we were able to find good correlation between HRV features and the emotion classes. Time domain features such as the mean heart rate, and the standard deviation of heart rate peaks were correlated with the emotion classes. These results show that with only 30 seconds of heart rate measurement, we were able to find features that efficiently help in the prediction of the elicited emotion. This finding could greatly benefit future research dealing with emotion recognition from physiological signals, as most previous research deal with long duration of heart rate measurements.

## 7. Conclusion

In this study we explored the effect of combining EEG and audio features on the classification accuracy of trained machine learning models to estimate the emotional state based on EEG data using the Emotiv Epoc headset. The categorical model of emotion was adopted compared with the dimensional model which is widely used. We used a film music dataset which was created to induce strong emotions and consisting of five basic emotions: Fear, angry, happy, tender and sad. EEG data was obtained from listening tests of labeled music excerpts. We extracted EEG and Audio features, and later applied machine learning techniques to study the improvement in the emotion recognition task using multimodal features. We evaluated learning curves of the mono and multimodal system to study the behavior of the machine learning systems for different training and test sizes, additionally we evaluated feature selection methods and their effect on the classification accuracies.

The results show that the multi-modality outperforms the EEG mono-modality. Preliminary results indicate a significant contribution of individual audio features in the classification accuracy, being consistent with previous research. The individual contribution of each audio features in the classification accuracy was studied. The inclusion of audio features with EEG features significantly increased the classification accuracy. Additionally, results are consistent with previous research dealing with the relationship between audio features and emotion ratings using the same dataset.

We also proposed a framework for the fusion of EEG and short term physiological signals for music emotion recognition. Preliminary results show the efficiency of using physiological signals such as skin conductance and heart rate variability features measure in a very short duration.

## Appendix

### Online resources

An online repository was created in order to share all of this thesis codes and data.

Link to github :

<https://github.com/JimmyJarjoura/Multimodal-classification>

## 8. List of figures

Figure 1: Available electrode positions on the Emotiv EPOC according to the 10-20 electrode placement system .....	15
Figure 2: BITalino biomedical data acquisition dev board .....	16
Figure 3: OpenSignals software for data acquisition, visualization and processing of physiological signals .....	17
Figure 4: The main components of skin conductance computed from the raw signal using decomposition analysis .....	24
Figure 5: Output analysis of the Kubios software of a sample, and all the features extracted.....	26
Figure 6: Comparison of knn classification accuracy for different k values and two distance metrics for three participants. ....	30
Figure 7: Comparison of EEG classification accuracy with and without feature selection for different training/testing size.....	35
Figure 8: Multimodal classification accuracy, before and after the inclusion of Audio features.....	37
Figure 9: Learning curve of multimodal and single-source systems for three participants. ....	36
Figure 10: EEG Classification accuracy amelioration after the inclusion of individual audio features.....	41

## 9. List of tables

Table 1: Overview of research dealing with emotion prediction from physiological signals .....	5
Table 2: Overview of research dealing with emotion classification from brain activity (EEG) .....	9
Table 3: EEG extracted features .....	20
Table 4: Features extracted from audio .....	22
Table 5: Skin conductance extracted features and their description .....	25
Table 6: EEG mood classification results. (v) Indicates that the accuracy using k-NN significantly outperformed SVM.....	29
Table 7: Comparison of EEG mood classification using all features with an approach using Wrapper feature selection. (v) indicates that the accuracy using wrapper feature selection significantly outperforms the classification using all features.....	31
Table 8: Multimodal (EEG and Audio) classification results for three participants (v) Indicates that the accuracy using the multimodal approach significantly outperforms the classification using EEG as a stand-alone method. ....	33
Table 9: Table comparing the results of our study with previous research dealing with correlation between audio features and emotion rating from participants .....	40
Table 10: Multimodal classification results (EEG and Skin conductance).....	41
Table 11: Correlation of ECG features with the emotion classes. ....	42

## References

- Aarts, L. a M. et al., 2014. A Multimodal Database for Affect Recognition and Implicit Tagging. *BMC medical imaging*, 3(1), p.9.
- Benedek, M. & Kaernbach, C., 2010a. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1), pp.80–91.
- Benedek, M. & Kaernbach, C., 2010b. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4), pp.647–658.
- Bos, D.O., 2006. EEG-based Emotion Recognition - The Influence of Visual and Auditory Stimuli. *Emotion*, 57(7), pp.1798–806.
- Castellano, G., Kessous, L. & Caridakis, G., 2008. Emotion recognition through multiple modalities: Face, body gesture, speech. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 92–103.
- Chanel, G. et al., 2009. Short-term emotion assessment in a recall paradigm. *International Journal of Human Computer Studies*, 67(8), pp.607–627.
- Chang, C. & Lin, C., 2013. LIBSVM : A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2, pp.1–39.
- Cheng, B. & Liu, G., 2008. Emotion Recognition from Surface EMG Signal Using Wavelet Transform and Neural Network. *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*, (1), pp.1363–1366.
- Choppin, A., 2000. EEG-Based Human Interface for Disabled Individuals : Emotion Expression with Neural Networks Submitted for the Master Degree. *Emotion*.
- Cover, T. & Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), pp.21–27.
- Daly, I. et al., 2015. Music-induced emotions can be predicted from a combination of brain activity and acoustic features. *Brain and Cognition*, 101, pp.1–11.
- Daly, I. et al., 2014. Neuroscience Letters Neural correlates of emotional responses to music : An EEG study. *Neuroscience Letters*, 573, pp.52–57.
- Daly, I. et al., 2012. What does clean EEG look like? In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. pp. 3963–3966.

- Delorme, A. & Makeig, S., 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), pp.9–21.
- Eaton, J., Williams, D. & Miranda, E., 2014. Affective Jukebox : a Confirmatory Study of Eeg Emotional Correlates in Response To Musical Stimuli. *Proceedings ICMC/SMC/2014*, (September), pp.580–585.
- Eerola, T. & Vuoskoski, J.K., 2011. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1), pp.18–49.
- Gabrielsson, A. & Lindström, E., 2010. The role of structure in the musical expression of emotions. In *Music and emotion: Theory, research, and applications*. pp. 367–400.
- Gosselin, N. et al., 2005. Impaired recognition of scary music following unilateral temporal lobe excision. *Brain*, 128(3), pp.628–640.
- Guerreiro, J. et al., 2013. BITalino: A Multimodal Platform for Physiological Computing. *Proc. of the 10th ICINCO Conf*, pp.500–506.
- Gunes, H. & Pantic, M., 2010. Automatic, Dimensional and Continuous Emotion Recognition. *International Journal of Synthetic Emotions*, 1(1), pp.68–99.
- Heraz, a, Razaki, R. & Frasson, C., 2007. Using machine learning to predict learner emotional state from brainwaves. *Advanced Learning Technologies 2007 ICAIT 2007 Seventh IEEE International Conference on*, 0(Table 1), pp.853–857.
- Hönig, F. et al., 2009. Classification of user states with physiological signals: On-line generic features vs. specialized feature sets. In *European Signal Processing Conference*. pp. 2357–2361.
- Ishino, K. & Hagiwara, M., 2003. A feeling estimation system using a simple electroencephalograph. *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, 5.
- Jerritta, S. et al., 2011. Physiological signals based human emotion Recognition: a review. *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*, pp.410–415.
- Juslin, P.N. & Laukka, P., 2004. Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. *Journal of New Music Research*,

- 33(3), pp.217–238.
- Juslin, P.N. & Sloboda, J.A., 2001. Communicating Emotion in Music Performance: A Review and Theoretical Framework. In *Music and Emotion*. pp. 310–331.
- Juslin, P.N. & Sloboda, J.A., 2010. *Handbook of Music and Emotion: Theory, Research, Applications*,
- Khalifa, S. et al., 2008. Positive and negative music recognition reveals a specialization of mesio-temporal structures in epileptic patients. *MUSIC PERCEPTION*, 25(4), pp.295–302.
- Kim, J., 2007. Bimodal emotion recognition using speech and physiological changes. *Robust Speech Recognition and Understanding*, (June), pp.265–280.
- Kim, J. & André, E., 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12), pp.2067–2083.
- Kim, Y.E. et al., 2010. Music Emotion Recognition : a State of the Art Review. *Information Retrieval*, (Ismir), pp.255–266.
- Kira, K. & Rendell, L.A., 1992. A practical approach to feature selection. In *In Proceedings of the ninth international workshop on Machine learning*. p. 249–256.
- Koelsch, S. et al., 2006. Investigating emotion with music: An fMRI study. *Human Brain Mapping*, 27(3), pp.239–250.
- Kononenko, I., 1994. Estimating attributes: Analysis and extensions of RELIEF. In Springer, Berlin, Heidelberg, pp. 171–182.
- Lang, P., Bradley, M. & Culthbert, B., 1999. International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2). *The Center for Research in Psychophysiology*,.
- Laurier, C., 2011. Automatic Classification of Musical Mood by Content-Based Analysis. *Group*, p.160.
- Laurier, C. et al., 2009. Exploring relationships between audio features and emotion in music. *Frontiers in Human Neuroscience*, 3(Escm), pp.260–264.
- Lee, N., Broderick, A.J. & Chamberlain, L., 2007. What is “neuromarketing”? A discussion and agenda for future research. *International Journal of Psychophysiology*, 63(2), pp.199–204.
- Levenson, R.W., 1988. Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity.
- Lin, Y.P. et al., 2010. EEG-based emotion recognition in music listening. *IEEE Transactions on*

- Biomedical Engineering*, 57(7), pp.1798–1806.
- Lin, Y.P. et al., 2007. Multilayer perceptron for EEG signal classification during listening to emotional music. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*.
- Makeig, S. et al., 2011. First demonstration of a musical emotion BCI. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 487–496.
- Müller, K.R. et al., 2008. Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167(1), pp.82–90.
- Niedermeyer, E. & Silva, F.H.L. Da, 2004. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*,
- Picard, R.W., 2003. Affective computing: Challenges. *International Journal of Human Computer Studies*, 59(1–2), pp.55–64.
- Ramirez, R. & Vamvakousis, Z., 2012. Detecting emotion from EEG signals using the Emotive Epoc device. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7670 LNAI, pp.175–184.
- Ricard, J., 2004. Towards computational morphological description of sound by. *Group*, (September).
- Sawata, R., Ogawa, T. & Haseyama, M., 2015. Human-centered favorite music estimation: EEG-based extraction of audio features reflecting individual preference. *International Conference on Digital Signal Processing, DSP*, 2015–Sept(July), pp.818–822.
- Schimmack, U. & Grob, A., 2000. Dimensional models of core affect: A quantitative comparison by means of structural equation modelling. *European Journal of Personality*, 14(March 1999), pp.325–345.
- Sethares, W.A., 2005. *Tuning, timbre, spectrum, scale: Second edition*,
- Shin, S. et al., 2014. MyMusicShuffler: Mood-based music recommendation with the practical usage of brainwave signals. *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, pp.355–356.
- Sourina, O., Liu, Y. & Nguyen, M.K., 2012. Real-time EEG-based emotion recognition for music therapy. *Journal on Multimodal User Interfaces*, 5(1–2), pp.27–35.
- Takahashi, K., 2004. REMARKS O N EMOTION RECOGNITION FROM MULTI-MODAL

- BIO-POTENTIAL SIGNALS. , pp.1138–1143.
- Tarvainen, M.P. et al., 2014. Kubios HRV - Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, 113(1), pp.210–220.
- Thammasan, N., Fukui, K. & Numao, M., 2016. Fusion of EEG and Musical Features in Continuous Music-emotion Recognition. , pp.1–8.
- Trochidis, K., Delbé, C. & Bigand, E., 2011. Investigation of the relationships between audio features and induced emotions in Contemporary Western music. *Proceedings of the 8th Sound and Music Computing Conference, SMC 2011*.
- Vempala, N.N. & Russo, F. a, 2012. Predicting Emotion from Music Audio Features Using Neural Networks. *International Symposium on Computer Music Modeling and Retrieval (CMMR)*, (June), pp.19–22.
- Wang, X.-W., Nie, D. & Lu, B.-L., 2014. Emotional state classification from EEG data using machine learning approach. *Neurocomputing*, 129(APRIL 2014), pp.94–106.
- Yurci, E., 2014. Emotion Detection From Eeg Signals : Correlating Cerebral Cortex Activity. *Department of Information and Communication Technologies,Universitat Pompeu Fabra, Barcelona*.
- Zentner, M. & Eerola, T., 2009. Self-report measures and models. In *Handbook of Music and Emotion*. pp. 187–221.
- Zhang, Q. & Lee, M., 2009. Analysis of positive and negative emotions in natural scene using brain activity and GIST. *Neurocomputing*, 72(4–6), pp.1302–1306.
- Zhu, X., 2010. Emotion Recognition of EMG Based on BP Neural Network. *Proceedings of the Second International Symposium on Networking and Network Security (ISNN '10)*, 1, pp.227–229.
- Zong, C.Z.C. & Chetouani, M., 2009. Hilbert-Huang transform based physiological signals analysis for emotion recognition. *Signal Processing and Information Technology (ISSPIT), 2009 IEEE International Symposium on*, pp.334–339.