Emotion Recognition based on EEG Signals in Response to Bilingual Music Tracks

Rida Zainab and Muhammad Majid
Department of Computer Engineering, University of Engineering and Technology Taxila, Pakistan

Abstract: *Emotions are vital for communication in daily life and their recognition is important in the field of artificial intelligence. Music help evoking human emotions and brain signals can effectively describe human emotions. This study utilized Electroencephalography (EEG) signals to recognize four different emotions namely happy, sad, anger, and relax in response to bilingual (English and Urdu) music. Five genres of English music (rap, rock, hip-hop, metal, and electronic) and five genres of Urdu music (ghazal, qawwali, famous, melodious, and patriotic) are used as an external stimulus. Twenty-seven participants consensually took part in this experiment and listened to three songs of two minutes each and also recorded self-assessments. Muse four-channel headband is used for EEG data recording that is commercially available. Frequency and time-domain features are fused to construct the hybrid feature vector that is further used by classifiers to recognize emotional response. It has been observed that hybrid features gave better results than individual domains while the most common and easily recognizable emotion is happy. Three classifiers namely Multilayer Perceptron (MLP), Random Forest, and Hyper Pipes have been used and the highest accuracy achieved is 83.95% with Hyper Pipes classification method.*

Keywords: *Emotion recognition, electroencephalography, feature extraction, classification, bilingual music.*

Received September 16, 2019; accepted July 26, 2020
https://doi.org/10.34028/iajit/18/3/4

1. Introduction

Affective computing is an interdisciplinary field that spans computer science, cognitive science, and psychology. It is based on the development of systems with the ability to detect, interpret, recognize, and simulate human affects. Emotion analysis has gained an important role in many industries today. Communication significantly relies upon emotions and one's ability to express them properly. Emotion recognition has its part in various fields like robotics, health, and eLearning [1, 9, 16]. Nowadays people have an inclination towards automation of various systems like smart homes, smart café's, and other such human and machine interaction-based systems. For such systems it is important to recognize user's emotions independently e.g., for the case of a smart home system may need to detect and recognize a person's mood to put on music and vary lights accordingly. So, knowledge of the user's emotions in a smart environment is vital and leads to effective results in terms of correctness and authenticity.

Emotion recognition has been a difficult problem because there exist different opinions by different researchers. William James tried to give a convincing answer back in 1884 i.e., "bodily changes following directly the perception of an exciting fact, and that our feeling of the same changes as they occur is the emotion" [40], but it started a continuing debate. Another definition, that describes emotion as "emotion is defined as an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism" [35]. Emotions can be categorized into aesthetic and utilitarian emotions. Utilitarian emotions correspond to emotions like anger, fear, joy, disgust, shame, sadness, and guilt. These emotions can cause one to change their behaviour towards events [34]. Emotions can be considered as positive and negative as well. Positive emotions include happiness, joy, pleasure, etc., whereas negative emotions consist of fear, disgust, anger, etc., [25].

Emotions can be represented either categorically or dimensionally. The categorial perspective says that there are some basic emotions and their combination results in other emotions. Different researchers have tried to define these basic emotions, Plutchik defines eight emotions as basic namely, joy, acceptance, surprise, curiosity, disgust, sadness, anger, and fear [33]. Another study has classified human emotions into six basic emotions i.e., anger, fear, disgust, sadness, surprise, and happiness [12]. On the other hand, dimensional perspective is based on cognition in which emotions are mapped to arousal, valence, and dominance dimensions. Valence scale represents a measure of positive feelings that goes from unpleasant to pleasure or sadness to happiness, arousal represents the level of excitement as in sleepy to excited and dominance shows the level of strength for particular emotion [23]. Most commonly the Circumplex model

is used that only incorporates valence and arousal scales.

There exist various techniques for human emotion recognition. Mainly speech, facial expression, and physiological signals-based modalities have been used for emotion recognition. All of these techniques have pros and cons associated with them. Speech and facial expressions are easier to collect using a microphone and camera respectively, but they suffer from biased opinions. It is easy for a user to hide emotion from speech or facial expression, which results in misclassification in speech and facial expression-based emotion recognition systems. Ali *et al*. [3], emotion recognition using facial images is proposed, which claims that people express emotions by facial expressions in the normal routine and showed better precision results, but the problem of authentic emotion recognition persists because facial expressions can be faked by individuals. Mirsamadi *et al*. [30], emotion recognition using a speech signal is proposed. Nasser and Sever [31], sentiment analysis based on language data using different classifier is presented. Speech data varies according to different subjects and makes the recognition process difficult but that is not the main problem in speech-based systems, we are unable to assure that emotion, correctly recognized by the system are real or not, meaning emotions expressed through speech can also be faked. Physiological measures such as sweat that can be gauged by Galvanic Skin Response (GSR), heart rate measure by PhotoplethysmoGraphy (PPG) or Electrocardiogram (ECG), and brain signals acquired using Electroencephalography (EEG) have been used for human emotion recognition [14, 28]. Emotions recognized through brain signals are more authenticated as they describe emotions directly from the human nervous system.

Signals produced in the human limbic system can give clues to cognitive activity in the brain, and these signals can be measured using EEG. EEG signals are basically used to measure the electrical activity of the human brain that changes with the change in brain functions like sleep, epilepsy, or any other activity. EEG signals can be divided into two types on the basis of recording from the scalp or intracranial [38]. Intracranial EEG signals are measured from inside the brain during brain surgery, while scalp EEG signals are measured by placing electrodes on the human scalp. EEG signals recorded from the human scalp are further categorized into two types i.e., monopolar and bipolar. Monopolar EEG recordings measure the potential difference between active electrodes and reference electrodes while bipolar EEG recordings measure the voltage difference between two active electrodes. Typically, EEG signals measured from the scalp have a range of 10-100 μV [41]. On the basis of frequency, EEG signals are divided into five bands namely gamma waves (>30Hz), beta waves (13-30Hz), alpha waves (8-13Hz), theta waves (4-7Hz), and delta waves (1-4Hz) [37].

EEG signals are effective for recognition of human emotions, induced when a proper stimulus is applied. The stimulus is anything that when presented to the subject, can evoke emotions. Emotion elicitation is possible in two ways i.e., subject elicited and event elicited. In the subject elicited case, participants are asked to remember some past emotional scenarios of their lives or try to feel a particular emotion. In the event elicited approach, a particular stimulus like photographs, music, videos, odour simulations, or tactile is presented to a participant for evoking emotions. Even using such stimuli, either subject's self-reported emotions are incorporated or standard stimulus sets of images and sounds such as International Affective Picture System (IAPS) [22], Geneva Affective Picture Database (GAPED) [11], and International Affective Digitized Sound System (IADS) [7] are used. In literature, different studies have used different stimuli that include images, audio clips, video clips, own memories, and games [2].

In the past few decades, many researchers have contributed to moving this system one step closer to perfection. Different studies have utilized different methods and stimuli to evoke and recognize human emotions using EEG signals [4, 15, 44]. EEG signals are recorded using various kinds of equipment either an EEG cap with electrodes on specified places or commercially available headsets like emotiv and muse headband. Raw EEG signals are not used to describe brain functionality rather salient features are extracted from EEG signals that can help to determine the state of brain activity. Generally, features are extracted in three different domains namely time, frequency, and wavelet. Wavelet features including relative wavelet energy, associative wavelet transform, and frequency domain features on individual bands were used to design an emotion recognition system using neural networks [20]. Candra *et al*. [8], a combination of wavelet entropy and an average of wavelet coefficients are explored as EEG features for classification of valence and arousal. In the time domain, power features are utilized for emotion recognition in an unsupervised learning scenario [21]. Power spectral entropy and correlation dimension are used to examine brain complexity for eight valence levels to recognize emotions [42]. Power Spectral Density (PSD) for all frequency bands is used as a feature set to classify emotions [17]. Mehmood *et al*. [27], six statistical features and frequency domain features are used to recognize four emotions namely happy, calm, scared, and sad.

For Brain-Computer Interface (BCI) based applications emotion analysis is conducted in [28] where four emotional responses (happy, sad, scared, and calm) are classified using Hjorth parameters. A study provided an evaluation method for the emotion

recognition system by exploiting the concept of frontal brain asymmetry [32]. Another study has used frequency domain features and three different classifiers namely K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and multilayer perceptron classifiers to recognize four emotions [43]. EEG and other physiological signals such as skin conductance and electrodermal activity have also been used to design multimodal emotion recognition systems as well [19, 45]. But in this paper, we focused on emotion recognition using a single modality i.e., EEG in response to music.

1.1. Related Work

In a recent study dynamical graph convolutional neural network is used to recognize positive, neutral, and negative emotions [39]. Subject dependent and subject independent classification methods have opted for two datasets namely Shanghai Jiao Tong University (SJTU) Emotion EEG Dataset (SEED) and Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices (DREAMER). Mert and Akan [29], the Database for Emotion Analysis using Physiological signals (DEAP) dataset is used for emotion recognition. Signals are decomposed into empirical oscillations i.e., intrinsic mode functions and Hjorth, correlation, and power-related features are used to recognize human emotions. DEAP and SEED datasets are used for two-state emotion recognition in [24]. They have used multiple features to find a generalization of features among different subjects and found Hjorth parameters to be more effective.

Another study has classified four emotions (happy, love, sad, and anger) and utilized time, frequency, and wavelet domain features [5]. Dataset was acquired using English audio music as stimuli for thirty subjects. Bo *et al*. [6], emotion recognition in response to audio music clips is proposed. They have analyzed the effects of music on human cognitive processes using EEG brain maps and observed that some music can give more arousal levels than others. They have utilized acoustic features of audio music signals, and PSDs of different frequency bands of EEG signals. Dataset used in [6] is self-acquired EEG signals recorded for each subject while listening to music. Nine-point SAM scale is used to determine the states of subjects. Ramirez *et al*. [35], music effects on cancer patients have been analyzed by recognizing their emotional responses. They concluded that music can have a positive effect on cancer patients and it also helps in relieving tiredness and anxiety.

Inspired from convolutional networks, a 3D convolutional network-based model for emotion recognition using EEG signals is proposed [13]. A benchmark DEAP dataset for emotion analysis is used and. data is augmented to have a 3D representation of the input. Data is classified into low/high classes for valance and arousal. A subject-independent emotion recognition system is proposed using convolution neural networks to recognize emotional states such as arousal and valence [18]. Dataset used by this study is self-acquired from 12 healthy male subjects using an EEG cap on selected set music audio.

As far as emotion recognition systems are concerned, it has been observed that most of the studies have used pre-recorded datasets and some have used their own recorded data in response to audio and video music in a single language. None of the above-mentioned studies have used bilingual music as a stimulus to evoke emotions. Our work is focused on the analysis of EEG signals in response to bilingual audio music tracks to recognize four human emotions.

1.2. Our Contribution

In this paper, EEG signals are recorded to recognize four different emotional states while the subject is presented with bilingual audio music tracks. Emotions are represented using dimensional valence arousal modal and characterize into four classes. Features are extracted from time and frequency domains; which are then utilized to recognize emotions. The supervised learning method is used for classification and three different classifiers namely Multi Layer Perceptron (MLP), Random Forest, and Hyper Pipes are used particularly. List of contributions of our work is;

1. A new dataset of EEG signals in response to bilingual music tracks is recorded using a four-channel headband that is commercially available and easy to wear.
2. Human emotional behavior in response to bilingual music is analyzed for the first time to the best of our knowledge.
3. Hybrid features of time and frequency domains are used to enhance the emotion recognition accuracy in response to bilingual music.

The rest of the paper is arranged as follows; Section 2 explains the methodology of the proposed emotion recognition system. Section 3 presents the experimental results followed by the conclusion in section 4.

2. Materials and Methods

In this study, bilingual audio music is presented to subjects in a noise-free environment while subjects wore EEG headband for signal acquisition. The methodology followed in this study is shown in Figure 1, which consists of data recording, pre-processing, feature extraction, and classification stages and are discussed in detail as follows;

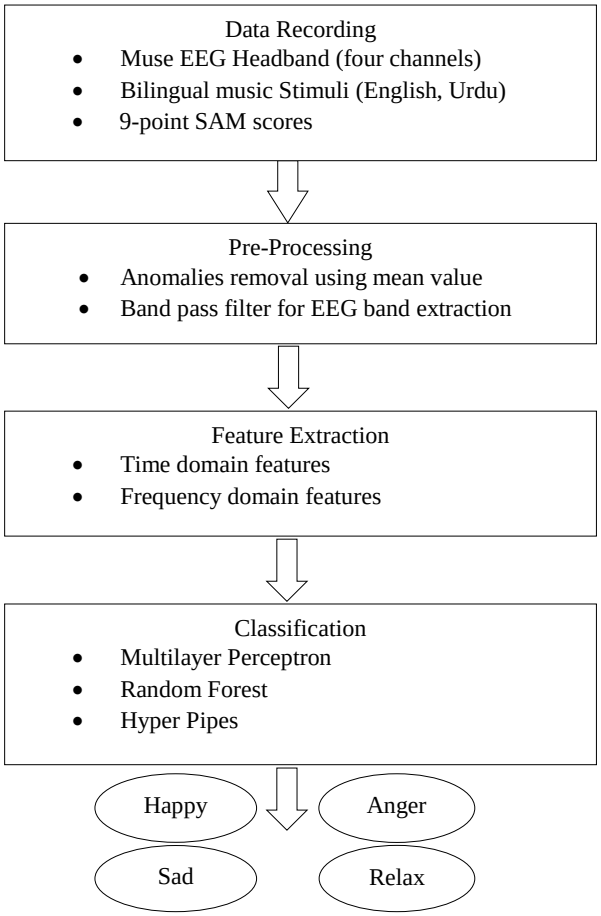

Figure 1. The proposed emotion recognition system using EEG signals in response to bilingual audio music.

2.1. Data Recording

EEG signals are acquired in a silent environment with minimal hustle, using a muse four-channel headband. Muse is a commercially available headband that is easy to handle. Subjects were briefly introduced to the purpose of the study and setup of the experiment. Consent paper was signed before data recording. A detailed view of the data recording setup is shown in Figure 2. Participants fill demographic details including age, gender, educational background, and music interests in the first phase of the experiment. In the second phase, subjects listen to three audio songs (each of two minutes), and their EEG data were recorded. After every song, an interface with SAM scale shown in Figure 3 was presented for twenty seconds to fill the valence and arousal scores. Valence score is the measure of happiness from gloomy to happy (1-9) and arousal score similarly represents the level of excitement (1-9). Other questions are asked as shown in the interface as well i.e., either you like the song or not and if you are already familiar with it or not.

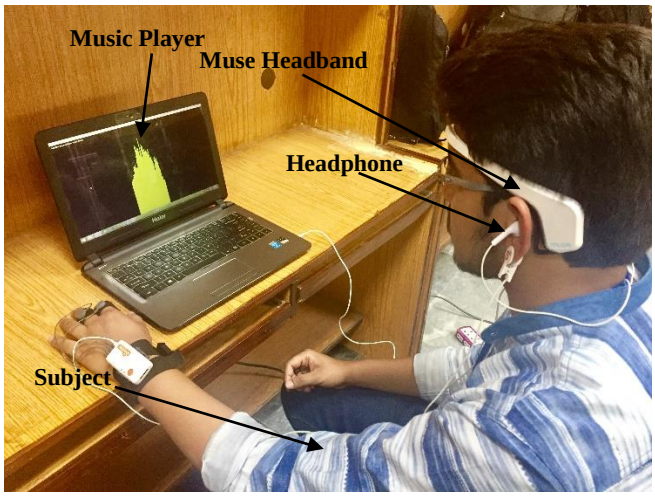

Figure 2. Experimental setup used in this study.

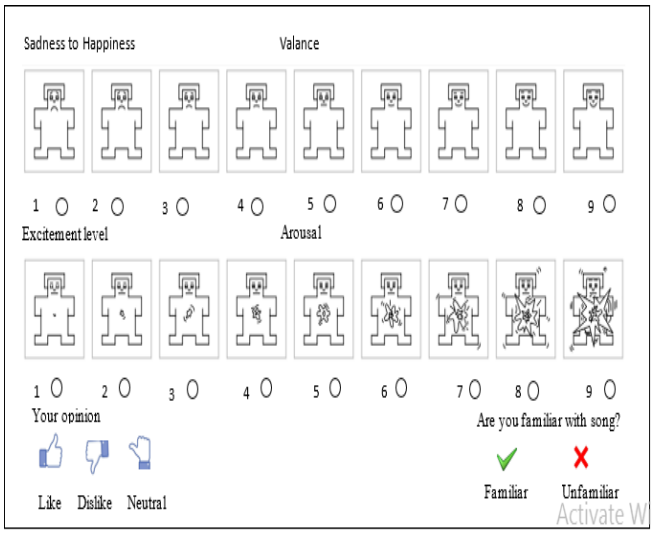

Figure 3. Nine point SAM scale interface for valence and arousal scores after listening to each song.

2.1.1. Participants

Twenty-seven healthy participants (13 males and 14 females) took part in the experiment with their consent. They belonged to three different age groups i.e., 18-25, 26-35, and 36-45 years and have different educational backgrounds ranging from bachelor students to Ph.D. faculty members. Subjects had different interests regarding music.

2.1.2. Stimuli

To evoke emotions, it is important to use certain stimulus that can help arise different emotions. In this experiment, bilingual audio music has been used for evoking emotions in participants. We have used five genres from English music (Electronic, Rap, HipHop, Metal, and Rock) and five genres from Urdu music (Patriotic, Melodious, Ghazal, Qawwali, and Famous).

2.1.3. Equipment

Equipment used in this study is a new and commercially available four-channel muse headband, which is easy to wear and adjustable to different skull

shapes and sizes. Muse is the most versatile and easy-to-use EEG system available. Electrodes of headband follow the 10/20 placement system (standard for electrode positioning). The four channels of muse headband are AF7, AF8, TP9, and TP10. The sampling rate of muse headband was set to 256Hz.

2.1.4. Dataset Description

The acquired data from all subjects comprised of eighty-one instances. Each participant is presented with three different songs from different genres from two different languages i.e., English and Urdu in a pseudo-random manner. EEG data of each subject in response to each song is stored in the form of an excel file where columns represent four channels and data type (raw data or band data) while rows represent data samples. The experiment lasted for about half an hour on average, but useful data was when the participant was listening to music, which consists of six minutes with 256 samples per second and two minutes for each song resulting in a total of 30720 samples for every occurrence.

2.2. Pre-Processing

Preprocessing is required when we have noisy or unwanted data. For reliable and accurate findings, it is vital to pre-process recorded data to remove any anomalies. In this work, preprocessing is performed by applying the mean function on raw data. The mean function calculates an average value for each column and corrupted data is replaced by that value. Five frequency bands i.e., delta, theta, alpha, beta, and gamma are extracted from each EEG channel data by applying an appropriate bandpass filter. The dataset acquired in this work has biased classes means an unequal number of instances for an individual class, which leads to false results as minority class is most of the time neglected by the classifiers. To mitigate this problem, the resampling technique is applied that introduces replicated instances to minority class, attempting to balance the dataset.

2.3. Feature Extraction

Brain signals provide tremendous information related to many events. To make EEG data useful for classification purposes, it is important to extract salient features that minimize computations and ease the process of further manipulation. In this work, features are extracted from time and frequency domains, which are explained in the following subsections.

2.3.1. Time Domain Features

a. Higuchi Fractal Dimension: It measures complexity of signal in the time domain and gives an idea of how wave pattern changes, by assessing the fluctuations of EEG signal. It is measured by creating k new time series. Let the given time series is $x[1], x[2], \ldots, x[N]$, then the new sub-series are as follows

$$X_k^l = \left\{ x[l], x[l+k], \cdots, x[l+\left\lfloor \frac{N-l}{k} \right\rfloor k] \right\}, \quad (1)$$

Where l=1, 2, 3, …k. and N is total number of samples. The length of each series is calculated as;

$$L_l(k) = \frac{1}{k} \left[\left(\sum_{i=1}^{\left\lfloor \frac{N-l}{k} \right\rfloor} |x[l+ik] - x[l+(i-1)k]| \right) . \left\lfloor \frac{N-l}{k} \right\rfloor \right] \quad (2)$$

Average of $L_l(k)$ over l, for all k gives another series $L(k)$ and Higuchi Fractal Dimension (HFD) is defined as best linear least squares fit to the curve with $\ln(1/k)$ on y-axis and $\ln(L(k))$ on x-axis.

b. Hjorth Parameters: Hjorth parameters has been used for emotion recognition previously [28]. These are namely three parameters activity, mobility, and complexity. If the given signal is represented by $x[n]$, where n=1, 2, 3, …, N with $p(x)$ as probability mass function, then the expected value of signal denoted by $E[x]$ is defined as;

$$E[X] = \sum_{x; p(x)>0} x p(x) \quad (3)$$

Similarly, the variance is computed as,

$$\text{var}(x) = E[X^2] - E[X]^2 \quad (4)$$

Then activity, mobility, and complexity are calculated as,

$$Activity = \text{var}(x) \quad (5)$$

$$Mobility = \sqrt{\frac{\text{var}\left(\frac{dx(n)}{dn}\right)}{\text{var}(x(n))}} \quad (6)$$

$$Complexity = \frac{\sqrt{\frac{\text{var}\left(\frac{d^2x(n)}{dn^2}\right)}{\text{var}\left(\frac{dx(n)}{dn}\right)}}}{Mobility} \quad (7)$$

Average or mean of time domain signal is calculated as,

$$\mu = \frac{1}{N} \sum_{n=1}^{N} x(n) \quad (8)$$

Standard Deviation of signal is calculated as,

$$\sigma^2 = \sqrt{\text{var}(x)} \quad (9)$$

c. Mean of First Difference of signal $x(n)$ is found by,

$$\Delta = \frac{1}{N-1} \sum_{n=1}^{N-1} x(n+1) - x(n) \quad (10)$$

First difference normalized is calculated by following equation,

$$\Delta_d = \frac{\Delta}{\sigma} \tag{11}$$

Energy of time domain signal is calculated using following equation,

$$Energy = \sum_{n=1}^{N} x(n).*x(n) \tag{12}$$

d. Entropy of signal gives measure of information present in it, if $x(n)$ is signal then entropy of signal denoted by S is as follows;

$$S = \sum_i p(x_i)\log_2(p(x_i)), \tag{13}$$

Where $p(x)$ is probability of signal.

e. Root mean square of signal $x(n)$ is calculated using following equation;

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N-1} x^2(n)} \tag{14}$$

f. Power of signal is calculated as;

$$P = \frac{1}{N}\sum_{n=1}^{N-1} x^2(n) \tag{15}$$

g. Kurtosis shows the sharpness of frequency distribution curve and it is as;

$$k = \frac{E(x-\mu)^4}{\sigma^4} \tag{16}$$

Where $E(x)^n$ represents nth moment of signal while μ and σ represents mean and variance of signal.

h. Minimum of time domain signal $x(n)$ is calculated by,

$$\min imum = \min\{x(n)\} \tag{17}$$

i. Maximum of time domain signal $x(n)$ is calculated by,

$$\max imum = \max\{x(n)\} \tag{18}$$

j. Skewness in terms of probability theory, is measure of asymmetry of data around sample mean. Skewness of signal $x(n)$, denoted by S is calculated as follows;

$$S = \frac{E(x-\mu)^3}{\sigma^3} \tag{19}$$

k. Signal peak to peak value is calculated as difference of maximum amplitude value and minimum amplitude value, denoted by *spp*;

$$spp = \max imum - \min imum \tag{20}$$

Peak to peak time window is calculated as difference of time when maximum and minimum amplitude values are spotted.

l. Signal peak to peak time slope denoted by *sppts* is calculated by dividing signal peak to peak value with its time window;

$$sppts = \frac{spp}{t_1 - t_2} \tag{21}$$

Where t_1 and t_2 are time when maximum and minimum amplitude values are spotted.

m. Correlation denoted by *corr* defines the measure of dependence between each of four channels of EEG data. If we consider each channel data as random variable with N scalar observations then it is computed as;

$$corr(A,B) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{A_i - \mu_A}{\sigma_A}\right)\left(\frac{B_i - \mu_B}{\sigma_B}\right) \tag{22}$$

2.3.2. Frequency Domain Features

Power, mean, standard deviation, first difference, and second difference are calculated for each individual band and all channels. Second difference is calculated as,

$$\Delta = \frac{1}{N-2}\sum_{n=1}^{N-2} x(n+2) - x(n), \tag{23}$$

Where N represents total number of samples.

Another feature calculated in frequency domain is Average of Square of Fourier Transform denoted by (ASFT) that is found by following equation;

$$ASTF = \frac{1}{N}\sum_{n=1}^{N}\frac{|F.*F|}{2T}, \tag{24}$$

Where F is the Fourier transform of time domain signal and T is the time period.

2.4. Classification

Classifying the given feature set into targeted classes is the final step. In this work, we have classified our feature set into four classes of emotion namely, happy, sad, relax, and anger. Classification is simply the identification of unknown objects from a class of known objects. Problem solvers attempt to solve this problem by finding a pattern and selects output based on pre-enumerated solutions, so output or prediction may not always be correct [10]. In this study, four emotion classes are taken and data is labelled based on self-reported emotions. Labelled data is fed to each classifier with default parameters, the tool used for classification is weka open source. We have used a 10-fold cross-validation technique for training and testing data split. Classifiers that have been used are MLP, Random Forest, and Hyper Pipes. Detail of each classifier is given below;

2.4.1. Multilayer Perceptron

MLP used a feed-forward neural network, which consists of input, output, and multiple hidden layers. It works by fine-tuning the weights to define the non-linear relationship between input and output. Neural networks perform classification in two steps i.e., training and testing. MLP works on the idea of backpropagation to train the network for sample data points. In the testing stage, a trained neural network is used to classify data [5].

2.4.2. Random Forest

It classifies data based on input from multiple trees. It constructs ensembles of decision trees during training phase to minimize the effect of overfitting. Each decision tree takes is assigned weight and gives its decision about the class of given data that in turn contributes to final decision. Overall decision is taken by considering output of all trees and then average result is final [36].

2.4.3. Hyper Pipes

This works by making a hyper pipe for each class, which consists of all points of that relevant category and records the attribute bounds observed for that class. It counts attributes (internally defined) for all samples and sample with the highest count is chosen. Classification is done on the basis of such counts and that is the reason why this classifier is fast. A study [26] has used this classifier for the emotional model using EEG signals.

3. Experimental Results

In this paper, we have presented the results in two parts i.e., statistical analysis of EEG bands in response to English and Urdu music and emotion recognition performance. The following subsections describe the experimental results in detail.

3.1. Statistical Analysis

Statistical analysis is performed to find out the relationship between different groups of music used in this study i.e., English and Urdu. Since stimulus used in this study is bilingual audio music tracks i.e., English and Urdu. Therefore, to identify the significant difference on arousal and valance scores of the subjects in response to English and Urdu music, a statistical t-test is applied. We divided our subjects into two groups i.e., one group that has listened to English music, while the second group listened to Urdu music. The groups having p-value less than 0.05 are considered as significantly different. The p-value obtained from the valence and arousal score in response to English and Urdu music are 0.03 and 0.04 respectively. This shows that the valence and arousal scores are significantly different in response to English and Urdu music tracks.

As EEG data is acquired from four channels namely TP9, AF7, AF8, and TP10. Even numbers and odd number electrodes represent right and left hemisphere of the brain respectively. We have to examine if left hemisphere and right hemisphere of brain function have significantly different in response to English and Urdu music. For this purpose, we have calculated Power Spectral Densities (PSD) for all four channels and applied t-test on PSDs of left hemisphere and right hemisphere in response to English and Urdu music. The p-value obtained from the PSDs from left and right hemispheres in response to English and Urdu music are 0.01 and 0.002 respectively. This shows that the left and right hemisphere activity are significantly different in response to English and Urdu music tracks.

3.2. Emotion Recognition Performance

Every participant listened to three songs from one of ten music genres from English and Urdu music tracks. For purpose of subjective evaluation, Self-Assessment Manikin (SAM) scale based perform was filled by the subject after listening to each music track. Responses recorded from SAM scale (1-9) are mapped to points -4 to 4 and plotted in valence-arousal plane as shown in Figure 4. All the points lying in first quadrant are labelled as happy, second quadrant is labelled as anger class with high arousal and low level of valence. Third quadrant is labelled with sad emotion class because it has low level of arousal and low level of valence. Similarly, points from fourth quadrant are represented with relaxed emotion class.

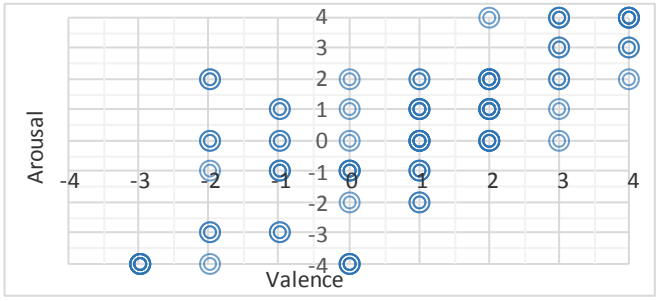

Figure 4. SAM scale scores plotted in valence-arousal plane.

Performance measures used in this study are average accuracy rate, F-measure, Root Means Squared Error (RMSE), and Mean Absolute Error (MAE). The accuracy of the classifier is the ratio of samples truly classified over total samples of data. F-measure gives information about the accuracy of the model with the added information of false positives and false negatives. Precision and recall are used to calculate F-measure, its value varies from 0 to 1, 1 representing the perfect F-measure while 0 represents worst. RMSE and MAE are measures of error performance of the classifier. RMSE shows prediction error as the difference between observed points and predicted points. Mean absolute error is the mean of prediction error for each instance of test data.

Three classifiers namely MLP, Random Forest, and Hyper pipes are used in this study for classification purposes. For classification results, a 10-fold cross-validation scheme is used. Data is split into equal-sized ten samples randomly, training, and testing sets. Nine splits are used to train the model while one testing split is used for validation. This process is repeated ten times while taking each of ten splits exactly once for validation. The final result is the average of all iterations.

For time-domain, features are calculated from all four channels of EEG data that result in a feature vector of length 108, which are further fed to classifiers for emotion recognition. Classifier results are shown in Table 1 for all the performance measures with the pre-processing step of resampling. For each classifier average classifier accuracy is tabulated, the hyper pipes classifier shows better results in terms of accuracy and F-measure. Since F-measure values are closer to 1 therefore, we have a better true positive rate as well as the rate of correctly predicted emotions. But on the other hand, prediction errors are higher as compared to MLP and Random Forest classifiers.

Table 1. Results of different classifiers against performance metrics using time domain features.

Classifier	Average accuracy %	F-measure	MAE	RMSE
MLP	70.37	0.694	0.1561	0.361
Random Forest	79.01	0.77	0.179	0.279
Hyper Pipes	80.24	0.795	0.359	0.415

Table 2. Confusion matrix for MLP classifier using time domain features with resample technique.

a	b	c	d	Classified as
37	2	0	0	a=Happy
6	12	2	0	b=Sad
2	1	13	0	c=Relax
4	0	0	2	d=Anger

Table 3. Confusion matrix for Random Forest classifier using time domain features.

a	b	c	d	Classified as
31	5	3	0	a=Happy
7	10	3	0	b=Sad
1	1	14	0	c=Relax
2	2	0	2	d=Anger

Table 4. Confusion matrix for Hyper Pipes classifier using time domain features.

a	b	c	d	Classified as
36	3	0	0	a=Happy
6	14	0	0	b=Sad
3	0	13	0	c=Relax
4	0	0	2	d=Anger

Tables 2, 3, and 4 present confusion matrices for MLP, Random Forest, and Hyper Pipe classifiers respectively using time-domain features. It can be observed that positive emotions like happiness is easier to recognize with higher true positive rates for all the three classifiers. Diagonal of all the confusion matrices represent true positives; this shows happy emotion is most of the time easily recognizable while anger emotion is difficult to recognize. As shown in Table 2 happy emotion has the largest value as true positives, a small number of false negatives i.e., 2 under column b. False positives are also greater in the case of a happy class. Similarly, in the case of the anger class, we have the lowest true positive rate or individual accuracy, but we also do not have any false positives. Relax class also has lower false positives. Hybrid features are a combination of time and frequency domain features that result in a larger feature vector of length 608. Classifier results are shown in Table 5 against all evaluation metrics using hybrid features. It is evident from results that by using hybrid features, classification accuracy is improved and Hyper Pipes classifier gives the best accuracy of 83.95% for the four-class problem.

Table 5. Results of different classifiers against performance metrics using hybrid features with resample technique.

Classifier	Average accuracy %	F-measure	MAE	RMSE
MLP	81.48	0.807	0.104	0.281
Random Forest	80.24	0.788	0.165	0.263
Hyper Pipes	83.95	0.835	0.344	0.400

Table 6. Confusion matrix of MLP classifier using hybrid features.

a	b	c	d	Classified as
37	1	1	0	a=Happy
3	13	1	3	b=Sad
2	0	14	0	c=Relax
2	2	0	2	d=Anger

Table 7. Confusion matrix of Random Forest classifier using hybrid features.

a	b	c	d	Classified as
37	1	1	0	a=Happy
6	12	2	0	b=Sad
2	0	14	0	c=Relax
3	1	0	2	d=Anger

Table 8. Confusion matrix of Hyper Pipes classifier using hybrid features.

a	b	c	d	Classified as
34	5	0	0	a=Happy
1	19	0	0	b=Sad
1	2	13	0	c=Relax
2	2	0	2	d=Anger

Table 9. Performance comparison of the proposed method with other methods.

Method	Stimulus	Subjects (F/M)	Emotion	Classifier	Acc %
Our	English and Urdu Music	27 (14/13)	Happy, Angry, Sad, Relaxed	Hyper Pipes	83.95%
[29]	Music Clips	32 (15/17)	Valence, Arousal	ANN	75.00%
[24]	Video Clips	15 (8/7)	Positive, Negative, Neutral	SVM	83.33%
[5]	Audio English Music	30 (15/15)	Happy, Love, Sad, Anger	MLP	78.11%
[6]	Music Clips	15 (7/8)	Valence, Arousal	SVM	66.80%
[13]	Music Videos	32 (15/17)	Valence, Arousal	CNN	88.49%
[18]	MIDI Songs	12 (0/12)	Valence, Arousal	CNN	86.87%

Confusion matrices of MLP, Random Forest, and Hyper Pipes classifiers using hybrid features are given in Tables 6, 7, and 8 respectively. Confusion matrices almost follow similar patterns as in the case of time-domain features. Happy emotion is still the most easily recognizable emotion. But now we have improved accuracy rates and the best accuracy achieved is 83.95% with Hyper Pipes classifier. False positives for a happy class are also decreased in number. Anger class has still lesser false positives as compared to others except for the case of MLP classifier as three instances of the sad class are

misclassified as anger emotion. The class imbalance could have affected the results because a happy class has a larger number of elements among other classes while on the other hand anger class consisted of the lowest number of instances. This led to false classification results. Resample technique help improve the recognition rate by balancing the number of instances between majority and minority classes, which resulted in improved accuracy rates. Time-domain results are satisfactory but when frequency domain-based feature vector is fused with a time-domain feature vector, results have improved greatly. Happy and sad emotions have better recognition rates as it can be seen from confusion matrices, relax and anger emotions are not well recognized as compared to happy and sad emotions, but their false positive rates are lower.

The proposed emotion recognition system is compared with the existing methods discussed in section 1.1. The comparison is performed in terms of stimuli used, the number of subjects involved in the study, the number of emotions to recognize, classifier used, and accuracy achieved and is shown in Table 9. The stimulus used in previous studies is music-based audio and emotional videos, whereas the proposed study used bilingual music as a stimulus for the first time. The proposed study has outperformed in case of average accuracy except [13, 18] because they have used CNN, which is time-consuming to train and required a large amount of data, whereas our method used handcrafted features, which is easy to train for the classifier.

4. Conclusions

Human emotions evoked in response to bilingual music are analysed and recognized in this work. Twenty-seven subjects listened to English and Urdu audio music tracks while wearing a muse headset for EEG data recording. Statistical analysis shows that English and Urdu music tracks are significantly different and evokes a different emotion. For emotion recognition, features from time and frequency domain are extracted and fed to three classifiers namely MLP, Random Forest, and Hyper Pipes. Hybrid features outperformed in general and the highest average accuracy of 83.95% is achieved by hyper pipes classifier. Happy and sad emotions are recognized with high recognition accuracies for all classifiers as compared to the anger and relaxed emotions.

References

[1] Akputu O., Seng K., Lee Y., and Ang L., "Emotion Recognition Using Multiple Kernel Learning Toward E-Learning Applications," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 14, no. 1, pp. 1-20, 2018.

[2] Alarcao S. and Fonseca M., "Emotions Recognition Using EEG Signals: A Survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374-393, 2017.

[3] Ali H., Hariharan M., Yaacob S., and Adom A., "Facial Emotion Recognition Using Empirical Mode Decomposition," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1261-1277, 2015.

[4] Atkinson J. and Campos D., "Improving BCI-Based Emotion Recognition by Combining EEG Feature Selection and Kernel Classifiers," *Expert Systems with Applications*, vol. 47, pp. 35-41, 2016.

[5] Bhatti A., Majid M., Anwar S., and Khan B., "Human Emotion Recognition and Analysis in Response to Audio Music Using Brain Signals," *Computers in Human Behavior*, vol. 65, pp. 267-275, 2016.

[6] Bo H., Ma L., Liu Q., Xu R., and Li H., "Music-Evoked Emotion Recognition Based on Cognitive Principles Inspired EEG Temporal and Spectral Features," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 9, pp. 2439-2448, 2019.

[7] Bradley M. and Lang P., "International Affective Digitized Sounds (IADS)," Technical Report, University of Florida, 1998.

[8] Candra H., Yuwono M., Chai R., Nguyen H., and Su S., "EEG Emotion Recognition Using Reduced Channel Wavelet Entropy and Average Wavelet Coefficient Features with Normal Mutual Information Method," *in Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society*, JeJu, pp. 463-466, 2017.

[9] Cavallo F., Semeraro F., Fiorini L., Magyar G., Sinčák P., and Dario P., "Emotion Modelling For Social Robotics Applications: A Review," *Journal of Bionic Engineering*, vol. 15, no. 2, pp. 185-203, 2018.

[10] Clancey W., *Classification Problem Solving*, Stanford University Stanford, 1984.

[11] Dan-Glauser E. and Scherer K., "The Geneva Affective Picture Database (GAPED): A New 730-Picture Database Focusing on Valence and Normative Significance," *Behavior Research Methods*, vol. 43, no. 2, pp. 468-477, 2011.

[12] Ekman P. and Davidson R., *The Nature of Emotion: Fundamental Questions*, Oxford University Press, 1994.

[13] Salama E., El-Khoribi R., Shoman M., and Shalaby M., "EEG-Based Emotion Recognition using 3D Convolutional Neural Networks," *International Journal of Advanced Computer*

Science and Applications, vol. 9, no. 8, pp. 329-337, 2018.

[14] Goshvarpour A., Abbasi A., and Goshvarpour A., "An Accurate Emotion Recognition System Using ECG and GSR Signals and Matching Pursuit Method," *Biomedical Journal*, vol. 40, no. 6, pp. 355-368, 2017.

[15] Hassan M., Alam M., Uddin M., Huda S., Almogren A., and Fortino G., "Human Emotion Recognition Using Deep Belief Network Architecture," *Information Fusion*, vol. 51, pp. 10-18, 2019.

[16] Huang H., Xie Q., Pan J., He Y., Wen Z., Yu R., and Li Y., "An EEG-Based Brain Computer Interface for Emotion Recognition and its Application in Patients with Disorder of Consciousness," *IEEE Transactions on Affective Computing*, pp. 1-1, 2019.

[17] Jirayucharoensak S., Pan-Ngum S., and Israsena P., "EEG-based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation," *The Scientific World Journal*, vol. 2014, 2014.

[18] Keelawat P., Thammasan N., Kijsirikul B., and Numao M., "Subject-Independent Emotion Recognition During Music Listening Based on EEG Using Deep Convolutional Neural Networks," *in Proceedings of the International Colloquium on Signal Processing and its Applications*, Penang, pp. 21-26, 2019.

[19] Kim K., Bang S., and Kim S., "Emotion Recognition System Using Short-Term Monitoring of Physiological Signals," *Medical and Biological Engineering and Computing*, vol. 42, no. 3, pp. 419-427, 2004.

[20] Krisnandhika B., Faqih A., Pumamasari P., and Kusumoputro B., "Emotion Recognition System Based on EEG Signals Using Relative Wavelet Energy Features and A Modified Radial Basis Function Neural Networks," *in Proceedings of the International Conference on Consumer Electronics and Devices*, London, pp. 50-54, 2017.

[21] Lan Z., Sourina O., Wang L., Scherer R., and Müller-Putz G., "Unsupervised Feature Learning for EEG-Based Emotion Recognition," *in Proceedings of the International Conference on Cyberworlds*, Chester, pp. 182-185, 2017.

[22] Lang P., "International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual," Technical Report, University of Florida, 2005.

[23] Lang P., "The Emotion Probe: Studies of Motivation and Attention," *American Psychologist*, vol. 50, no. 5, pp. 372, 1995.

[24] Li X., Song D., Zhang P., Zhang Y., Hou Y., and Hu B., "Exploring EEG Features in Cross-Subject Emotion Recognition," *Frontiers in Neuroscience*, vol. 12, pp. 162, 2018.

[25] MacIntyre P. and Vincze L., "Positive and Negative Emotions Underlie Motivation for L2 Learning," *Studies in Second Language Learning and Teaching*, vol. 7, no. 1, pp. 61-88, 2017.

[26] Mahfuz N., Ismail W., Jali Z., Anuar K., and Nordin M., "Classification of Brainwave using Data Mining in Producing an Emotional Model," *Journal of Theoretical and Applied Information Technology*, vol. 75, no. 2, pp. 128-136, 2015.

[27] Mehmood R. and Lee H., "A Novel Feature Extraction Method Based on Late Positive Potential for Emotion Recognition in Human Brain Signal Patterns," *Computers and Electrical Engineering*, vol. 53, pp. 444-457, 2016.

[28] Mehmood R., Du R., and Lee H., "Optimal Feature Selection and Deep Learning Ensembles Method for Emotion Recognition from Human Brain EEG Sensors," *IEEE Access*, vol. 5, pp. 14797-14806, 2017.

[29] Mert A. and Akan A., "Emotion Recognition from EEG Signals by Using Multivariate Empirical Mode Decomposition," *Pattern Analysis and Applications*, vol. 21, no. 1, pp. 81-89, 2018.

[30] Mirsamadi S., Barsoum E., and Zhang C., "Automatic Speech Emotion Recognition Using Recurrent Neural Networks with Local Attention," *in Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, New Orleans, pp. 2227-2231, 2017.

[31] Nasser A. and Sever H., "A Concept-based Sentiment Analysis Approach for Arabic," *The International Arab Journal of Information Technology*, vol. 17, no. 5, pp. 778-788, 2020.

[32] Petrantonakis P. and Hadjileontiadis L., "A Novel Emotion Elicitation Index Using Frontal Brain Asymmetry for Enhanced EEG-Based Emotion Recognition," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 5, pp. 737-746, 2011.

[33] Plutchik R., "The Nature of Emotions: Human Emotions Have Deep Evolutionary Roots, A Fact That May Explain Their Complexity and Provide Tools for Clinical Practice," *American Scientist*, vol. 89, no. 4, pp. 344-350, 2001.

[34] Qayyum H., Majid M., Anwar S., and Khan B., "Facial Expression Recognition using Stationary Wavelet Transform Features," *Mathematical Problems in Engineering*, vol. 2017, 2017.

[35] Ramirez R., Planas J., Escude N., Mercade J., and Farriols C., "EEG-Based Analysis of The Emotional Effect of Music Therapy on Palliative Care Cancer Patients," *Frontiers in Psychology*, vol. 9, pp. 254, 2018.

[36] Ramzan M. and Dawn S., "Learning Based Classification of Valence Emotion from Electroencephalography (EEG)," *International Journal of Neuroscience*, vol. 129, no. 11, pp. 1085-1093, 2019.
[37] Schomer D. and Da Silva F., *Niedermeyer 's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Lippincott Williams and Wilkins, 2012.
[38] Siuly S., Li Y., and Zhang Y., "EEG Signal Analysis and Classification," *IEEE Transactions on Neural Systems and Rehabilitaiton Engineering*, vol. 11, pp. 141-144, 2016.
[39] Song T., Zheng W., Song P., and Cui Z., "EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532-541, 2018.
[40] Takahashi K., "Remarks on Emotion Recognition From Bio-Potential Signals," *in Proceedings of the International Conference on Autonomous Robots and Agents*, Palmerston North, pp. 186-191, 2004.
[41] Teplan M., "Fundamentals of EEG Measurement," *Measurement Science Review*, vol. 2, no. 2, pp. 1-11, 2002.
[42] Tong J., Liu S., Ke Y., Gu B., He F., Wan B., and Ming D., "EEG-based Emotion Recognition Using Nonlinear Feature," *in Proceedings of the International Conference on Awareness Science and Technology*, Taichung, pp. 55-59, 2017.
[43] Wang X., Nie D., and Lu B., "EEG-based Emotion Recognition Using Frequency Domain Features and Support Vector Machines," *in Proceddings of the International Conference on Neural Information Processing*, Shanghai, pp. 734-743, 2011.
[44] Xu Q., Zhou H., Wang Y., and Huang J., "Fuzzy Support Vector Machine for Classification of EEG Signals Using Wavelet-Based Features," *Medical Engineering and Physics*, vol. 31, no. 7, pp. 858-865, 2009.
[45] Yoo G., Seo S., Hong S., and Kim H., "Emotion Extraction Based on Multi Bio-Signal Using Back-Propagation Neural Network," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4925-4937, 2018.

Rida Zainab completed her BSc. in Computer Engineering from University of Engineering and Technology (UET) Taxila, Pakistan in 2017 and received gold medal for her best performance. She is currently doing her MSc in Computer Engineering from UET Taxila, Pakistan.

Muhammad Majid received BSc. in Computer Engineering with honors from University of Engineering and Technology (UET) Taxila, Pakistan in 2005, MSc. in Data Communications with distinction and PhD in Electronic and Electrical Engineering from the University of Sheffield, UK in 2007 and 2011 respectively. He is currently an Associate Professor at Department of Computer Engineering, UET Taxila. His research interests include video coding and emotion recognition.