21st Iranian Conference on Biomedical Engineering (ICBME 2014), Biomedical Engineering Faculty, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, Nov 26-28, 2014

Recognition of Emotional States Induced by Music Videos Based on Nonlinear Feature Extraction and SOM Classification

S.Hatamikia[*1]
[*1]Department of biomedical engineering
Science and Research branch, Islamic Azad University
Tehran, Iran
[*1]shatamikia11@gmail.com

A.M Nasrabadi[2]
[2]Department of biomedical engineering
Shahed University
Tehran, Iran
[2]nasrabadi@shahed.ac.ir

Abstract—**This research aims at investigating the relationship between Electroencephalogram (EEG) signals and human emotional states. A subject-independent emotion recognition system is proposed using EEG signals collected during emotional audio-visual inductions to classify different classes of continuous valence-arousal model. First, four feature extraction methods based on Approximate Entropy, Spectral entropy, Katz's fractal dimension and Petrosian's fractal dimension were used; then, a two-stage feature selection method based on Dunn index and Sequential forward feature selection algorithm (SFS) algorithm was used to select the most informative feature subsets. Self-Organization Map (SOM) classifier was used to classify different emotional classes with the use of 5-fold cross-validation. The best results were achieved using combination of all features by average accuracies of %68.92 and %71.25 for two classes of valence and arousal, respectively. Furthermore, a hierarchical model which was constructed of two classifiers was used for classifying 4 emotional classes of valence and arousal levels and the average accuracy of %55.15 was achieved.**

Keywords: Emotion recognition; Nonlinear analysis; Sequential forward feature selection algorithm (SFS); Dunn index; Self Organization Map (SOM)

I. INTRODUCTION

Emotions have a key role in managing daily life of an individual and directly affect cognition, perception, attention, reasoning, decision making and memory. Daily activities are completely influenced by the individual's feelings and people need to express different states of their emotions to communicate with each other [1]. Emotion detection is one of the vital steps towards emotional intelligence in advanced Brain computer interfaces (BCIs); where, in such systems detecting some of emotions like stress and fear have vital role for patients who use these systems, in dangerous situations. According to neurophysiological researches, there is a relationship between physiological activities and emotions [2]. Hence, in several studies, different physiological signals such as respiration rate, heart rate, skin conductance, electromyogram and electrocardiogram are used to recognition of different emotional states [3, 4]. Furthermore, some studies have been suggested emotion recognition systems using speech signals and facial expressions [5]. Currently, most of researches are focused on EEG-based emotion recognition. To recognition of EEG correlates on emotional states, distinct emotions have to be defined. Six basic emotions, suggested by Ekman et al, are commonly accepted across cultures and nationalities. These emotions are including: sad, happy, surprise, fear, anger, and disgust. Moreover, some complex emotions are made from a combination of these basic emotions. In order to overcome this issue, a two-dimensional model is defined which can provide a continuous representation of emotions. This continuous model which is suggested by Russell measures emotions along two axes of valence and arousal dimension. Valence axis measures unpleasant to pleasant and arousal axis measures calm to excited. Several emotions can be simply mapped onto the valence–arousal space. Four main emotional classes formed by valence-arousal space are including: high arousal/high valence (HAHV), high arousal/low valence (HALV), low arousal/high valence (LAHV), low arousal/low valence (LALV). Fig.1 shows the distribution of different emotions in two dimensional valence/arousal space. Murugappan et al. [6] suggested an emotion recognition method using statistic features extracted from different EEG frequency bands and wavelet transform. They classified five discrete emotions (happy, surprise, fear, disgust, and neutral) with the average accuracy of %79.17. In another research of this authors, several discrete emotions were recognized using energy-based wavelet feature extraction and the average accuracy of %83.26 was achieved [7]. Soleymani et al. [8] proposed an emotion recognition system based on PSD features using pupillary response, gaze distance and EEG signals for classifying three classes of valence and arousal dimension. They achieved the best average accuracies of 68.5 and 76.4 for two and three classes of arousal and valence classes, respectively. Chung and Yoon [9] proposed a system based on power spectral density (PSD) features for classifying two and three classes of valence and arousal dimensions. They achieved the accuracy of 66.6% and 53.4% for two and three levels of valence dimension, respectively. Bastos Filho et al [10] suggested an emotion recognition system to discriminate three emotional states: calm, stress, and normal. They used three feature extraction methods to recognize different emotions from EEG signals. The feature extraction methods were based on PSD, statistical and high order crossing (HOC) features and finally, the best accuracy of %70.1 was achieved using PSD features. According to the literature review, in most studies the feature extraction method is based on power spectral features and wavelet transform. This present paper tries to evaluate

efficiency of nonlinear feature extraction methods in recognition of emotional states. For this aim, two nonlinear set of features used include Fractal-based feature extraction (Kat'z and Petrosian's method) and Entropy–based feature extraction (Approximate Entropy and Spectral Entropy). Then, the performance of these features was evaluated using a two-stage feature selection method and SOM classifier. The rest of this study is summarized as follows. Section II explains about data acquisition and subjects. Section III illustrates the research methodology include: feature extraction, feature selection and classification algorithm. Section IV presents the results and discussion and conclusion of this present research are given in section V.

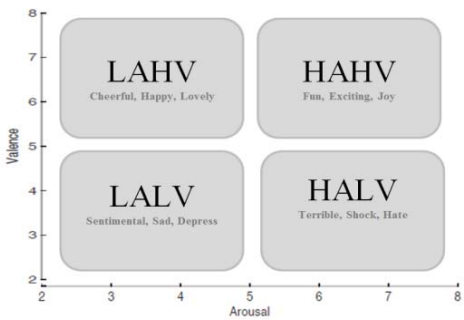

Figure1. Some example emotions in four emotional quadrants (HAHV, HALV, LALV, LALV)

II. DATA COLLECTION

In this study, a public data for emotion analysis using physiological signals (DEAP) is used which is recorded by Koelstra et al. [11]. This data include EEG and peripheral signals from 32 participants: 16 men and 16 women (average age of 26.9). EEG signals from 32 electrodes from different positions include: Fp1, AF3, F3, F7, FC5, FC1, C3, T7, CP5, CP1, P3, P7, PO3, O1, Oz, Pz, Fp2, AF4, Fz, F4, F8, FC6, FC2, Cz, C4, T8, CP6, CP2, P4, P8, PO4, and O2 were recorded according to 10-20 standard system. Koelstra et al. prepared the raw data and the preprocessed data and for each participant. In this research, we used the preprocessed EEG signals to evaluate our proposed method. In the preprocessed data, EEG signals were filtered (between 4.0 and 45 Hz) and were down sampled (at 128 Hz). Furthermore, eye artifacts were rejected by blind source separation method. 40 one min music videos were represented for each subject and the level of valence and arousal dimensions of participant's feeling was ranged using self-assessment manikins (SAM) questionnaire. SAM questionnaire was used to visualize the range of valence and arousal dimensions by manikins. Participants were asked to choice one number from 1 to 9 which was written below the manikins as represent in Figure 2.

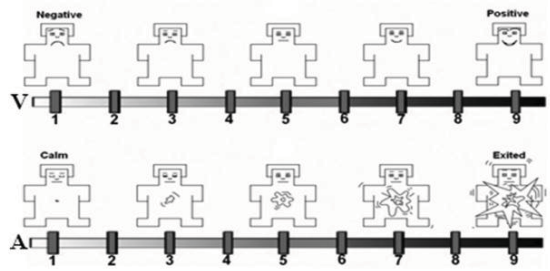

Figure 2. SAM questionnaire and degrees of valence (above) and arousal (bellow) dimension

III. METHODOLOGY

A. Feature Extraction

1) Katz's fractal dimension

The concept of fractal dimension refers to a non-integer dimension emanates from fractal geometry. Fractal dimension can be measured as the irregularity of a signal and has a direct dependence with the amount information inside the signals.
The Kat's fractal dimension of a signal is defined as follow:

$$D = \frac{\log_{10}(L/a)}{\log_{10}(d/a)} = \frac{\log_{10}(n)}{\log_{10}(d/L) + \log_{10}(n)} \qquad (1)$$

Where, d is the distance between the first sample of the time series and the sample that prepares the furthest distance. a and L are the average and sum of distances of sequential data points. Also, n is L/a.

2) Petrosian's fractal dimension

Petrosian's technique provide a quick estimate of the fractal dimension as follow:

$$D = \frac{\log_{10}(n)}{\log_{10}(n) + \log_{10}(\frac{n}{n + 0.4 * N})} \qquad (2)$$

Where, length of the time sequence is n and the number of sign variations in derivative of the signal is shown by N.

3) Approximate entropy

Approximate Entropy (ApEn) is a measure which can provide a description of the irregularity of the time series. The predictability of fluctuations in a series can be measured by this index. The value of ApEn would be larger when a complex process is less predictable.
At first, to compute the ApEn of time series x_i, $i=1,\ldots, n$, the state vectors in the embedding space should be constructed as follow:

$$y_i = \{x_i, x_{i+1}, x_{i+2,} \ldots, x_{(i+m-1)}\}, \quad 1 \le i \le n-m+1 \qquad (3)$$

Where, m is the embedding dimension. Second, $C_i^m(r)$ is defined as follow:

$$C_i^m(r) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} \theta(r - d(y_i, y_j)) \qquad (4)$$

Where, $\theta(y)$ is the standard heavyside function that $\theta(y)=1$ for y>1 and $\theta(y)=0$, otherwise. Also, r is the vector comparison distance. Distance measure $d(y_i, y_j)$ is defined as follow:

$$d(y_i, y_j) = \max_{k=1,2,\ldots,m} \left(\left| y_{(i+k-1)} - y_{(j+k-1)} \right| \right) \quad (5)$$

third, $\varphi^m(r)$ is calculated by:

$$\phi^m(r) = \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \ln C_i^m(r) \quad (6)$$

At the end, for fix n, m and r, approximate entropy A_e can be computed as follow:

$$A_e(m,r,n) = \phi^m(r) - \phi^{m+1}(r) \quad (7)$$

4) Spectral entropy

The Spectral Entropy (SE) related to frequency range $f1 - f2$ is compute as follow:

$$SE = \sum_{f_i=f_1}^{f_2} p(f_i) \log\left(\frac{1}{p(f_i)}\right) \quad (8)$$

Where, $P(fi)$ is the normalized PSD at frequency fi . The SE value is logarithm of the total number of frequency components in the range $f1 - f2$.

B. Feature Selection

- *Feature evaluation based on Dunn index*

Dunn index is a measured to identify compact and well separated clusters. It is a ratio-type criterion where the separation is measured by the maximum cluster diameter and the cohesion is measured by the nearest neighbor distance. The Dunn's measure D is defined as:

$$D = \min_{1 \le i \le k} \left[\min_{i+1 \le j \le k} \left[\frac{dist(c_i, c_j)}{\max_{1 \le l \le k} diam(c_l)} \right] \right] \quad (9)$$

Where, the distance between clusters c_i and c_j is $dist(c_i,c_j)$ and is computed as follow:

$$dist(c_i, c_j) = \min_{x_i \in c_i, x_j \in c_j} d(x_i, x_j) \quad (10)$$

Where, the distance between data points $x_i \epsilon ci$ and $x_j \epsilon c_j$ is $d(x_i,x_j)$. Also, $diam(c_l)$ is the diameter of cluster c_l and is defined as follow:

$$diam(c_l) = \max_{x_{l1}, x_{l2} \in c_l} d(x_{l1}, x_{l2}) \quad (11)$$

- *Feature subset selection based on sequential feature selection algorithm*

Sequential forward feature selection algorithm (SFS) algorithm is one of the most common feature subset selection methods. To obtain the best feature subset, this search algorithm subsequently adds to the first set of features which is considered empty at the first level of the procedure. At first, this algorithm seeks one of the features with the most influence in increasing the fitness function and adds the feature with the highest fitness; then, it seeks for the second feature that combination of it with the first selected feature lead to the best fitness function. This process continues until adding a new feature does not improve the performance.

C. Classification with Self-organization map

The self-organization map (SOM) algorithm is one of the most distinguished unsupervised method which is commonly used in time series prediction, vector quantization, and clustering. Moreover, this network is used in classification issues in various fields. In SOM network, a vector of weights w_i specified for each neuron i. The dimension of these weight vectors is considered equal to the dimension of the input data. Firstly, according to the input data $x(t)$, one neuron $i^*(t)$ is selected as bellow:

$$i^*(t) = \arg\min\{g_i(x(t))\},$$
$$where, g_i(x(t)) = \|x(t) - w_i(t)\| \quad (11)$$

Afterward, the weight vectors should be updated according to follow equation:

$$\Delta w_i(t) = \alpha(t) h(i*, i; t)\left[x(t) - w_i(t)\right] \quad (12)$$

Where, $h(i^*, i; t)$ is neighborhood function which is computed as:

$$h(i*, i; t) = \exp\left(-\left\| r_i(t) - r_{i^*}(t) \right\|^2 / \sigma^2(t)\right. \quad (13)$$

In the output layer, the distance between the location of the neurons i and i^* is $\|r_i(t)-r_{i*}(t)\|$. *Also,* $\sigma(t)$ and $\alpha(t)$ are the neighborhood radius and the learning rate parameters, respectively. To classification based on SOM, output neurons should be labeled. This procedure is done after the training phase and is called the labeling phase. First, the winner neuron is selected for each one of the training vectors; then, the class labels of the training vectors those select neuron i as the winner neuron are specified. At the end, the label of i_{th} winner neuron is determined according to the most frequent class labels of these training vectors. In the testing stage, the class labels of each new data is found by class label of it's selected winner neuron. In current research, the initial neighborhood size is set at 3 and after 100 steps it reaches to 1. Also, the initial $\alpha(t)$ is set at 0.8. SOM network was run 3 times to ensure the reproducibility of classification performances and the average accuracies over three runs were considered.

IV. EXPERIMENTAL RESULTS

In this research, EEG signals from 32 participants collected during emotional audio-visual inductions were used. Due to nonlinear nature and complex dynamics of EEG signals, we decided to use nonlinear methods to evaluate emotional states correlate to EEG signals. For this aim, two nonlinear set of features used include Fractal-based feature extraction (Katz's and Petrosian's method) and Entropy–based feature extraction (Approximate entropy and Spectral entropy). As we said earlier, 32 participants participated in the experiments and 40 one min music videos were represented for each subject; so, 1280(=32×40) instances were used. EEG signal of each instance were include nearly 8000 sample which were divided into 1000 sample segments.

The mentioned features were extracted from all segments. Furthermore, these features were extracted from all (32) channels. Therefore, for each feature type, the feature vector was constructed of 256(=8(segments) ×32(channels)) features. We used a two-stage feature selection shame based on Dunn criterion and SFS method to select the most informative

features and also to avoid of redundant features. At the first stage of the feature selection method, the Dunn index was calculated for each independent feature. Then, all the features were ranked according to their values of Dunn index and the number of 50 features with the maximum values of Dunn index were selected. At the second stage, the selected features of the first stage were used as input of the SFS algorithm to select the most discriminating subsets of feature. Accuracy of SOM classifier was used as fitness function of SFS algorithm. The performance of classification process was evaluated using K-fold cross validation with K=5. In our paper, emotional states were divided into two classes of valence and arousal levels according to the participant's SAM ratings. If rating of SAM questionnaire was larger than 5, the class label of valence and arousal was considered as high valence and high arousal, respectively. On the other hand, If rating of SAM questionnaire was smaller than 5, the class label of valence and arousal was considered as low valence and high arousal, respectively. F1-score and accuracy was used to evaluate performance of emotion classification .F1-score is a statistic measure that takes the class-balance into account. This measure is a harmonic mean of recall (the rate of correct results in true data) and precision (the rate of correct results in the classification results) which is defined as follow:

$$F1-score=\frac{2\times precision\times recall}{precision+recall} \quad (14)$$

The average accuracies over all participants and F1-score (average F1-score for both classes) for different feature types and all features are shown at Table I. As we have shown in Table I, the best accuracies among different features for both valence and arousal classes were achieved using Approximate entropy feature with the accuracy of %63.34 and %66.49 for valence and arousal classes, respectively. The results also showed entropy-based features had better performances as compare to fractal-based features. Furthermore, according to Table I, using all feature types together has improved classification results as compare to use of only one feature type. The achieved accuracies using all features are %68.92 and %71.25 for valence and arousal, respectively. We also proposed a hierarchical model to classify 4 emotional classes (HAHV, HALV, LAHV, LALV). The structure of proposed hierarchical model is shown at Figure 3. The hierarchical model was constructed of two classifiers of the same type that one classifier was trained to classify valence levels and the other was trained to classify arousal levels. SOM was used as classifier in proposed hierarchical model. The average accuracies and F1-scores for different classes of the hierarchical model are given in Table II.

TABLE I. AVERAGE ACCURACIES (ACC) OVER PARTICIPANTS AND F1-SCORES

Features	Valence		Arousal	
	ACC	F1 -score	ACC	F1-score
ApEn	%63.34	%51.92	%66.49	%51.25
SE	%61.40	%46.83	%65.82	%48.92
Katz	%58.73	%47.50	%62.57	%47.60
Petrosian	%59.42	%45.18	%60.23	%49.10
All features	%68.92	%49.69	%71.25	%50.84

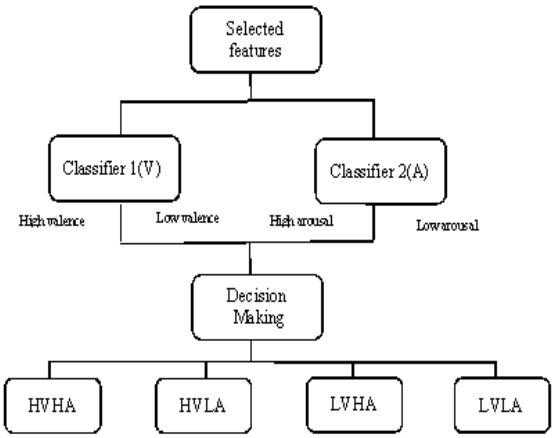

Figure 3. Structure of hierarchical model to classify 4 classes of valence-arousal model

TABLE II. AVERAGE ACCURACY AND F1-SCORE FOR FOUR CLASS OF VALENCE-AROUSAL

Class	HVHA	HVLA	LVHA	LVLA
F1-score	%71.20	%52.99	%57.98	%33.81
accuracy	%55.15			

V. DISCUSSION AND CONCLUSION

In this study, we examined two different types of nonlinear features: Entropy-based features and Fractal-based features to classify emotional states. We used SOM classifier for classification of emotions. This classifier performs locally on the data space and can prepare a good estimation of the input space; so, this network can be suitable for distinguishing different data structures with variant distributions. According to the results, our subject-independent method based on nonlinear features and SOM classifier is an efficient method for classifying different classes of continuous valence-arousal model. We found that Entropy-based features have better performances than Fractal-based features in discrimination of both valence and arousal dimension. Also, we found that combination of all feature types has increased classification accuracies by almost %5-%9 as compare to use of only one feature type. Such result is earned from this fact that variant features can represent the signal in different domains and more useful information could be provided using various feature types. Until now, almost in all studies using DEAP database, which has suggested subject-independent emotion recognition system with valence-arousal model, the feature extraction

method is based on power spectral features. Our proposed method based on nonlinear features have shown better results as compare similar studies using same database based on power spectral features[9,11,13]. We proposed a hierarchical model to classify 4 emotional states of valence and arousal model and achieved the average accuracy of %55.14; while, according to previous study using this database, the highest achieved accuracy for 3 classes of valence and arousal are %55.4 and %55.2, respectively [13].

There is no study on classifying 4 classes of valence and arousal using this database, and due to variety in data acquisition conditions, types of emotional states, number of participant and decision-making scheme, it is not possible to compare the results with other studies. According to the results of Table 3, the F1-score percentages for different emotional classes have shown a considerable differences where, the highest F1-score was obtained for class HAHV and the lowest F1-score was obtained for LALV class. From these results we can conclude that the classification of HAHV is more accurate than other classes and the classification of LALV is less accurate than other classes. It seems the reason is imbalance data. Since, according to SAM questionnaire, the number of instances which correlate to HAHV class are larger than the number of instances which associate to LALV class. However, for using the proposed system in real situations the classification accuracies must be increased higher. In future, we will examine our proposed method with other nonlinear features, feature selection and classification methods to improve the performance of emotion recognition system.

REFRENCES

[1] R. Picard R, E. Vyzas, J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state", IEEE Trans Pattern Anal Mach Intell, vol,23, pp.1175-1191, 2001.

[2] F. Nasoz , CL. Lisetti, K. Alvarez, N. Finkelstein, "Emotion recognition from physiological signals for user modeling of affect", 9th International Conference on User Model, pp.22-26, Pittsburg, USA, June 2003.

[3] JT. Cacioppo , LG. Tassinary, " Inferring psychological significance from physiological signals", Psychol, vol.45, pp.16-28,1990.

[4] M. Naji, M. Firoozabadi ,P. Azadfallah, "Classification of Music-Induced Emotions Based on Information Fusion of Forehead Biosignals and Electrocardiogram", Cogn Comput, vol.6, pp.241-252, 2014.

[5] R. Cowie , ED. Cowie, N. Tsapatsoulis, G Votsis, S. Kollias, W.Fellenz, et al,"Emotion recognition in human-computer interaction", IEEE Signal Process, vol.18, pp.32-28, 2001.

[6] M. Murugappan, R. Nagarajan, S. Yaacob, "Comparison of different wavelet features from EEG signals for classifying human emotions", IEEE Symposium on Industrial Electronics and Applications (ISIEA), Kuala Lumpur, Malaysia, pp. 4-6,Oct 2009.

[7] M. Murugappan, N. Ramachandran, Y. Sazali Y, "Classification of human emotion from EEG using discrete wavelet transform", J Biomed Sci Eng, vol.3, pp.390-396, 2010.

[8] M. Soleymani, M. Pantic, T. Pun, "Multimodal emotion recognition in response to videos. IEEE Trans Affect Comput ",vol.3, pp.211-223, 2012.

[9] SY. Chung, HJ, "Yoon Affective classification using Bayesian classifier and supervised learning", 12th International Conference on Control, Automation and System (ICCAS), Island, pp. 1768–1771, Oct 2012.

[10] TB. Filho , A. Ferreira, A. Atencio , S. Arjunan , D. Kumar," Evaluation of feature extraction techniques in emotional state recognition",IEEE Proceedings of 4th International Conference on Intelligent Human Computer Interaction, Kharagpur, India, pp.27-29, December 2012.

[11] S. Koelstra, C. Muhl, M. Soleymani, JS. Lee, A. Yazdani, T. Ebrahimi, et al,"IEEE transactions on affective computing. DEAP: A database for emotion analysis Using Physiol Signals", vol.3, pp.18-31, 2012.

[12] J.C. Dunn, "Well separated clusters and optimal fuzzy partitions", J. Cybernet, vol.4, pp. 95–104, 1974.

[13] HJ. Yoon, SY, " Chung. EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm", Comput Biol Med, vol.43, pp. 2230-2237, 2013.