Proceedings of the 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 11-13 November 2012

Emotion Classification Based on Forehead Biosignals using Support Vector Machines in Music Listening

Mohsen Naji
Department of Biomedical Engineering
Islamic Azad University, Science and Research branch
Tehran, Iran
m.naji@srbiau.ac.ir

Mohammad Firoozabadi
Department of Medical Physics
Tarbiat Modares University
Tehran, Iran
pourmir@modares.ac.ir

Parviz Azadfallah
Department of Psychology
Tarbiat Modares University
Tehran, Iran
azadfa_p@modares.ac.ir

Abstract—**The purpose of this study was to investigate the feasibility of using forehead biosignals as informative channels for classification of music-induced emotions. Classification of four emotional states in Arousal-Valence space was performed by employing two parallel support vector machines as arousal and valence classifiers. Relative powers of EEG sub-bands, spectral entropy, mean power frequency, and higher order crossings were extracted from each of the three forehead data channels: left Temporalis, Frontalis, and right Temporalis. The inputs of the classifiers were obtained by a feature selection algorithm based on a fuzzy-rough model. The averaged subject-independent classification accuracy of 93.80%, 92.43%, and 86.67% for arousal classification, valence classification, and classification of four emotional states in Arousal-Valence space, respectively, is achieved.**

Keywords-forehead biosignals; emotion classification; arousal; valence.

I. Introduction

Everyone knows that listening to music improve negative moods. Accordingly, music therapy has received increased attention over the last few years. Music therapists work with a variety of emotional and psychophysiological symptoms. Hence, it would be great if the therapists knew what effects a music excerpt has on human's emotional state.

Many studies have explored the effects of different kinds of music stimuli on central nervous system (CNS) and peripheral nervous system (PNS) responses. The research on CNS responses has shown that depending on the type of musical stimuli spectral power of EEG bands can be altered [1]; and frontal regions and auditory cortex appear to have certain activity for musical processing [2].

In recent years a number of researchers have been worked on the issue of emotion recognition during music listening. Kim and André [3] investigated the potential of surface electromyogram (EMG), electrocardiogram (ECG), Galvanic skin response (GSR), and respiration changes for emotion recognition during listening to music. Lin *et al.* [4] applied support vector machine (SVM) to categorize EEG dynamics according subject self-reported emotional states during music listening. However, the accuracy of emotion recognition systems are improving, many users may not feel comfortable with interfaces such as EEG-caps or respiration belts.

In order to providing user-friendly human-machine-interface (HMI) Firoozabadi *et al.* [5] proposed a novel biosignal acquisition method by locating three pairs of electrodes on participant's Frontalis and Temporalis forehead muscles. These forehead biosignals (FBS) convey both the information of their adjacent standard EEG locations as well as facial expression information. More recently, Rezazadeh *et al.* [6] applied the FBS signals to design a control interface which could be adapted to user's affective state.

The present paper, with the knowledge that the frontal regions have a key role in emotional processing, was designed to study the feasibility of using FBS signals for emotion classification during music listening. Classification of four musical emotions in the Arousal-Valence emotional space is performed by using features of FBS data. We utilize a feature reduction algorithm based on a generalized fuzzy-rough and employ the support vector machine to classify FBS signals.

II. Methods

First of all, emotional biosignals should be recorded by applying appropriate musical stimuli. After signal acquisition and preprocessing stage, we calculated four sort of temporal and spectral features from each of the biosignals and selected the most significant feature subsets and classifiers for arousal and valence classifications. Putting together the outputs of the valence and arousal classifiers, the four-class classification can be performed. In the following we explain the mentioned items in more detail

A. Emotional Stimuli

We place listeners' emotional states in the four quadrants of Arousal-Valence plane. Arousal, which varying from low to high, describes the extent of calmness or excitation felt by

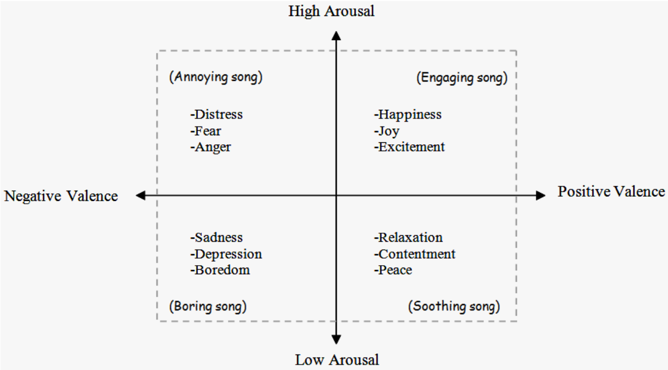

Fig. 1. Arousal-Valence model of emotion with example emotions.

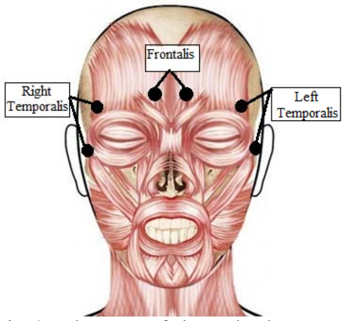

Fig. 2. The areas of electrode placement.

people; and valence, which varying from negative to positive, describes the level of pleasure or aversiveness. Hence, emotions can correspond to arousal-valence using a set of musically descriptive adjectives (Fig. 1): soothing (low arousal-positive valence), engaging (high arousal-positive valence), annoying (high arousal-negative valence), and boring (low arousal-negative valence). The music excerpts were chosen on the basis of self-ratings in a pilot study of 50 participants about 15 music pieces. The participants were asked to complete a questionnaire which consisted of questions about subjective feelings elicited by each music piece. After collecting the questionnaires the selection of music pieces was performed. The music excerpts were: Pachelbel's *Canon in D major* (Lee Galloway), a *Persian 6/8* song, *Hi friend!* (Deadmau5 featuring Mc Flipside), and *Romance* (Schumann's Symphony No. 4) for classes soothing, engaging, annoying, and boring, respectively.

B. Experiment procedure

Twenty-two healthy non-musician volunteers participated; all were right-handed subjects (7 males, and 15 females self-selected as not having premenstrual syndrome) in the age group of 20-23 years. None of the subjects had hearing impairment or history of mental disorders.

Before placing any electrode, the electrode placement area was abraded lightly and cleansed with alcohol to lower the electrical impedance. Three pairs of pre-gelled Ag/AgCl electrodes were placed on subject's facial muscles in a differentiation configuration to acquire FBS signals. One pair is places on subject's Frontalis muscle: above the eyebrows with about 3 cm inter-electrodes distance (Frontalis channel). Two pairs are places on left and right Temporalis muscles with about 4 cm inter-electrodes distance (left Temporalis channel and right Temporalis channel). Fig. 2 shows the location of electrode pairs. The ground electrode for FBS acquisition is placed on the left earlobe. The BIOPAC MP100 system (ack100w software version) was used to collect biosignals. The sampling frequency and amplifier gain were selected at 256 Hz and 5000, respectively.

Subjects were seated in a comfortable chair in a quiet non-sound proofed room with minimal light for the recording of the biosignals. They were instructed to keep their eyes closed, put on the headphones, and remain seated in the music-listening experiment. The data were recorded during 60-s silent followed by 120-s moderate intensity music. In order to validate the success of emotion induction the subjects were requested to complete a questionnaire to report their musical induced emotions after each run.

C. Preprocessing and feature extraction

Raw FBS signals were filtered (band-pass: 1-100 Hz; band-stop: 47-53 Hz). Then, all the filtered signals were divided into unequal portions: rest signal, which lasts 1 minute and acquired during silent; and emotional signal, which lasts 2 minutes and obtained during listening to a music excerpt. Emotional FBS data were divided into 30-s segments and in each segment 5 groups of features were computed. To nullify the resting pattern, the features were also computed for rest signals to calculate the normalized features, $F_{normal}(j)$, as follow:

$$F_{normal}(j) = \frac{F_{emotional}(j) - F_{rest}}{F_{rest}}, \qquad j = 1,2,3,4 \qquad (1)$$

where F_{rest} is a feature value extracted from rest signal, and $F_{emotional}(j)$ is a feature value extracted from j^{th} 30-s segment of emotional signal. These normalized features were applied for classifications. The features used to form observation matrix are defined as given in the following.

Relative powers

After estimation of power spectral density of the signals from fast Fourier transform, relative powers (RP) [1] were calculated for following frequency bands: theta (θ: 4-7 Hz), slow alpha (α_1: 8-10 Hz), fast alpha (α_2: 11-13 Hz), alpha (α : 8-13 Hz), slow beta (β_1: 13-19 Hz), fast beta (β_2: 20-30 Hz), beta (β: 13-30 Hz), and gamma (γ: 31-50 Hz). The relative powers were calculated as total power of a frequency band divided by signal power.

Spectral entropy

The spectral entropy [7] corresponding to the frequency range $f_1 - f_2$ is formulated as:

$$S = \sum_{f_i=f_1}^{f_2} P(f_i) \log\left(\frac{1}{P(f_i)}\right), \qquad (2)$$

where $P(f_i)$ is normalized PSD at frequency f_i. Thereafter, the entropy value is divided by logarithm of the total number of frequency components in the range $f_1 - f_2$. In this paper the frequency range of 4-35 Hz was selected.

Mean frequency

The mean frequency (MF) [8] in the range $f_1 - f_2$ is calculated as

$$MF = \frac{\sum_{f_i=f_1}^{f_2} f_i \times P(f_i)}{\sum_{f_i=f_1}^{f_2} P(f_i)}, \qquad (3)$$

where $P(f_i)$ is PSD at frequency f_i. We selected the range of 4-35 Hz for MF calculation.

Higher Order Crossings (HOC)

Higher order crossings (HOC) [9] are obtained by counting the number of zero-crossings in the filtered time series. The HOC of order m, HOC_m, for a zero-mean time series of $x(n)$ can be calculated as

$$HOC_m = \mathrm{NZC}\{\nabla^{m-1}(x(n))\}, \qquad m = 1,2,3,\dots \qquad (4)$$

where ∇ is backward difference operator; and NZC{.} denotes the number of zero crossings. In this paper, HOC_1 to HOC_8 were calculated and divided by duration of the time series.

D. Feature evaluation criterion

To evaluate the extracted features, we used a novel feature significance measure, presented by Hu *et al.* [10], which derived from a generalized fuzzy-rough model. After normalizing the observations of each dimension in the feature space, the significance measure will be obtain after some calculations given in the following.

In a feature space of *F*, the fuzzy equivalence class $[x_i]_F$ of observation x_i is defined as

$$[x_\mathrm{i}]_F = \frac{r_{i1}}{x_1} + \frac{r_{i2}}{x_2} + \cdots + \frac{r_{iN}}{x_N}, \qquad (5)$$

where *N* is the total number of observations, "+" means the union, and r_{ij} is the output of a symmetrical membership function which measures the value of the fuzzy similarity degree between x_i and x_j. That is, $r_{ij} = f(|x_i - x_j|)$, where $|x_i - x_j|$ means Euclidean distance between x_i and x_j. In this paper, a gaussian similarity relation function was adopted:

$$r_{ij} = \exp\left(-\left(x_i - x_j\right)^2/2\sigma^2\right), \sigma = 0.25 \ . \qquad (6)$$

We can define the lower approximation of the decision *X* as

$$\underline{F_k}X = \{x_i | I([x_i]_F, X) \geq k, x_i \in U\}, \qquad 1 \geq k \geq 0.5, \qquad (7)$$

where $I(A,B) = \frac{\sum_{x\in U}\mu_{A\cap B}(\mathrm{x})}{\sum_{x\in U}\mu_{\mathrm{A}}(\mathrm{x})}$, and $\mu_{\mathrm{A}}(\mathrm{x})$ is membership degree of x in the fuzzy set A. *k* is a parameter which reflects users' tolerance degrees of noise. The less the *k*, the more the users can tolerance noise. For a two-class problem, the lower approximation of classification *D* is defined as

$$\underline{F_k}D = \{\underline{F_k}X_1, \underline{F_k}X_2\}. \qquad (8)$$

Finally, the feature significance measure of feature space *F* for classification *D* is calculated as

$$\gamma = \frac{|\underline{F_k}D|}{N}, \qquad (9)$$

where |.| is the cardinality (number of elements) of a set. Obviously, $0 \leq \gamma \leq 1$. The greater the γ, the higher class separability.

E. Feature selection method

For each parameter *k* of the described feature evaluation criterion, we used the sequential forward floating selection (SFFS) approach [11] to select informative feature subsets, *F*. Thereafter, the feature subsets were imposed to classifiers and classification rates were calculated.

F. Classifiers

This study employed support vector machines (SVM) [11] for pattern classification. In the SVM classifiers, a radial basis function is used as a kernel function for data projection. To design a FBS-based emotion recognition system which classifies the music-induced emotions to four classes of positive valence-low arousal, positive valence-high arousal, negative valence-high arousal, and negative valence-low arousal we considered the following steps to be performed:

- Putting the extracted data in two classes of high arousal and low arousal based on their labels.
- Changing the threshold *k* of the feature evaluation criterion from 0.5 to 0.95 with step 0.05; selection of features for high arousal-low arousal classification problem; and determining average arousal-classification accuracy for each *k*.
- Determining the threshold k_A, for the most accurate arousal classifier.
- Putting the extracted data in two classes of positive valence and negative valence based on their labels.
- Changing the threshold k of the feature evaluation criterion from 0.5 to 0.95 with step 0.05; selection of features for positive valence-negative valence classification problem; and determining average valence-classification accuracy for each *k*.
- Determining the threshold k_V, for the most accurate valence classifier.
- Putting together the outputs of k_A-based arousal classifier and k_V-based valence classifier to develop FBS-based emotion recognition system.

In order to estimate the true power of the arousal or valence classifiers, a twenty five times of 4-fold cross validation technique was adopted.

III. Results

A. ArousalClassification

According to Table I, arousal classification accuracy and selected features of each channel vary with the specified threshold of the feature evaluation criterion. We can find that for k_A=0.75 the average accuracy rate is maximum (93.8%). The selected feature subset, labeled as ADFS (arousal discriminant feature subset), for classifying high arousal and low arousal emotions is made from concatenation of the following features: $RP_{\alpha 1}$ and RP_γ of left Temporalis channel, $RP_{\alpha 1}$,RP_β, and SE of Frontalis channel, and HOC_2, HOC_3, and HOC_5 of right Temporalis channel.

TABLE I

SELECTED FEATURES AND CORRESPONDING AROUSAL CLASSIFICATION RATES VERSUS THE PARAMETER OF THE FEATURE EVALUATION CRITERION

k	Selected features for left Temporalis channel	Selected features for Frontalis channel	Selected features for right Temporalis channel	Classification Accuracy
0.5	RP_{γ}	HOC_7		65.06 ±6.97 %
0.55	RP_{θ}, RP_{β}	RP_{θ}, HOC_2	HOC_3	88.74± 5.21 %
0.6	$RP_{\beta1}$, HOC_8	HOC_3	HOC_3, HOC_7	86.22± 5.68 %
0.65	$RP_{\beta1}$, RP_{γ}, MF, HOC_8		HOC_2, HOC_5	88.71± 4.56 %
0.7	RP_{θ}, RP_{γ}	$RP_{\alpha1}$, $RP_{\alpha1}$, RP_{β}	HOC_3, HOC_5	92.34±4.12 %
0.75	$RP_{\alpha1}$, RP_{γ}	$RP_{\alpha1}$,RP_{β}, SE	HOC_2, HOC_3, HOC_5	**93.80± 3.47** %
0.8	RP_{γ}, MF	RP_{θ}, $RP_{\alpha1}$, $RP_{\beta1}$, $RP_{\beta2}$	$RP_{\alpha2}$, $RP_{\beta1}$, HOC_3	89.88± 4.85 %
0.85	RP_{γ}, MF, HOC_1, HOC_8	RP_{θ}, $RP_{\alpha1}$, $RP_{\alpha2}$, $RP_{\beta1}$	RP_{θ}, $RP_{\alpha2}$, HOC_3	91.11± 4.28 %
0.9	RP_{θ} ,$RP_{\alpha1}$, HOC_1	RP_{θ} ,$RP_{\alpha1}$, $RP_{\beta1}$, HOC_7	MF, HOC_1, HOC_2	90.14± 4.96 %
0.95	SE, HOC_3, HOC_8	RP_{α}		63.97± 7.52 %

k: The parameter of fuzzy-rough model (feature evaluation criterion).
The bolded rates indicate the most classification accuracies among the all feature subsets (p<0.05).

TABLE II

SELECTED FEATURES AND CORRESPONDING VALENCE CLASSIFICATION RATES VERSUS THE PARAMETER OF THE FEATURE EVALUATION CRITERION

k	Selected features for left Temporalis channel	Selected features for Frontalis channel	Selected features for right Temporalis channel	Classification Accuracy
0.5	RP_{θ}			69.88± 6.31 %
0.55	RP_{θ}			69.88± 6.52 %
0.6	$RP_{\alpha1}$			70.00± 6.34 %
0.65	$RP_{\beta1}$			69.74± 5.99 %
0.7			SE	70.68± 7.23 %
0.75	HOC_1	RP_{θ} ,$RP_{\alpha1}$, $RP_{\beta1}$, HOC_7	$RP_{\alpha1}$, HOC_1	**92.06± 4.29** %
0.8	HOC_1	RP_{θ}, $RP_{\alpha1}$, $RP_{\beta1}$, RP_{β}	$RP_{\alpha1}$, HOC_1, HOC_2, HOC_8	**92.43± 4.42 %**
0.85	RP_{θ}, RP_{γ}, HOC_1	RP_{θ}, $RP_{\alpha1}$, $RP_{\alpha2}$, $RP_{\beta1}$	$RP_{\alpha1}$, MF, HOC_3, HOC_8	**91.74± 4.77** %
0.9	RP_{θ}, MF, HOC_1	RP_{θ}, RP_{α}, $RP_{\beta1}$, HOC_5	RP_{θ}, HOC_1, HOC_2, HOC_8	90.48± 4.74 %
0.95	RP_{θ}, $RP_{\beta1}$, HOC_1, HOC_7	RP_{θ}, $RP_{\alpha1}$, $RP_{\alpha2}$, $RP_{\beta1}$, HOC_1	RP_{θ}, MF, HOC_1, HOC_4	**91.68± 4.34** %

k: The parameter of fuzzy-rough model (feature evaluation criterion).
The bolded rates indicate the most classification accuracies among the all feature subsets (p<0.05).

B. Valence Classification

Table II presents that valence classification accuracy and selected features of each channel vary with the specified threshold of the feature evaluation criterion. We can find that four set of features, labeled as VDFS1, VDFS2, VDFS3, and VDFS4 (valence discriminant feature subset), significantly obtain the highest mean classification accuracy (92.06%, 92.43%, 91.74%, and 91.68% corresponding to k_V=0.75, 0.8, 0.85, and 0.95) among selected feature subsets. However, the maximum mean classification accuracy for classifying positive valence and negative valence emotions is obtained from concatenation of the following features: HOC_1 of left Temporalis channel, RP_{θ} , $RP_{\alpha1}$, $RP_{\beta1}$, and RP_{β} of Frontalis channel, and $RP_{\alpha1}$, HOC_1 , HOC_2 , and HOC_8 of right Temporalis channel.

C. Final Results

The results of arousal and valence classifications reveal that the emotions which induced by listening to music can be classify to high arousal or low arousal, and positive valence or negative valence by using two parallel SVM classifier.

Applying an input pattern, the output 0 or 1 is obtained for each classifier. Therefore, a four class emotion recognition system can be designed by combining the arousal and valence classifiers.

The overall classification accuracies and for the different optimum feature subsets are tabulated in Table III. The maximum (not significantly) overall classification accuracy is obtained 86.67±4.83 % by putting together the outputs of the ADFS-input arousal classifier and VDFS2-input valence classifier.

TABLE III

OVERALL CLASSIFICATION RATES AND SENSITIVITY VALUES FOR THE SELECTED FEATURE SUBSETS

Feature Type of Arousal Classifier	Feature Type of Valence Classifier	Classification Accuracy (%)
ADFS (k_A=0.75)	VDFS1 (k_V=0.75)	86.39±5.70
ADFS (k_A=0.75)	VDFS2 (k_V=0.80)	86.67±4.83
ADFS (k_A=0.75)	VDFS3 (k_V=0.85)	86.05±5.52
ADFS (k_A=0.75)	VDFS4 (k_V=0.95)	86.01±5.29

IV. DISCUSSION

The present study was performed to demonstrate the feasibility of using forehead biosignals for classifying music-induced emotions. After acquisition and preprocessing of the FBS signals, the following features were extracted: relative powers (RP) of EEG spectrum sub-bands, spectral entropy (SE), mean frequency (MF), and higher order crossings. Two SVM classifiers were designed separately and combined afterwards: arousal classifier, and valence classifier. The inputs of the classifiers varied according to parameter *k* of the feature evaluation algorithm.

As be seen in Table I and Table II, RP features and HOC features were dominantly selected for arousal or valence classification. The selected features support the conclusions of previous researches which revealed that there are associations between brain wave powers and emotional states [1]. Besides, HOC features involve in formation of feature subsets. HOC features were recently introduced and utilized for EEG-based emotion recognition [9]. Our results verify the effectiveness of these features for FBS data classification.

The best arousal classification rate, the best valence classification rate, and corresponding total classification rate were obtained 93.80±3.47%, 92.43±4.42%, and 86.67±4.83%, respectively. Currently, researchers are working on developing emotion recognition systems, especially by applying musical stimuli. Using four-channel biosignals Kim and André [3] presented an average classification rate of 70% for subject-independent PNS-based emotion recognition across 4 subjects. Lin *et al.* [4] proposed an EEG-based emotion recognition system for distinguishing 4 musical induced emotions with the maximum average accuracy of 82.29±3.06% across 26 subjects. Compared with previously published works on emotion recognition during music listening, we achieved higher classification rates. Moreover, the proposed FBS-based classification has the advantages of subject independency, and small number of data channels. By using the proposed emotion recognition system, we hope to see increasing progress in the fields of music therapy and interactive multimedia systems.

ACKNOWLEDGMENT

We gratefully acknowledge the assistance of Ms Atena Bajoulvand for her help with collection of the data of female subjects.

REFERENCES

[1] E. M. Sokhadze, “Effects of music on the recovery of autonomic and electrocortical activity after stress induced by aversive visual stimuli,” *Appl Psychophysiol Biofeedback*, vol. 31, no. 1, pp. 31-50, Mar 2007.

[2] L. A. Schmidt, and L. J. Trainor, “Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions,” *Cognition and Emotion*, vol. 15, pp. 487-500, 2001.

[3] J. Kim, and E. André, “Emotion Recognition Based on Physiological Changes in Music Listening,” *IEEE Trans Pattern Anal Mach Intell*, vol. 30, no. 12, pp. 2067-2083, Dec 2008.

[4] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. H. Chen, “EEG-Based Emotion Recognition in Music Listening,” vol. 57, no. 7, pp. 1798-1806, Jul 2010.

[5] S. M. P. Firoozabadi, M. R. A. Oskoei, and H. Hu, “A Human–Computer interface based on forehead Multi-Channel bio-signals to control a virtual wheelchair,” in *Proc. 14th Iranian Conference on Biomedical Engineering*, Tehran, Iran, 2008, pp. 108-113.

[6] I. M. Rezazadeh, X. Wang, M. Firoozabadi, M. R. H. Golpayegani, “Using affective human– machine interface to increase the operation performance in virtual construction crane training system: A novel approach,” *Automation in Construction*, vol. 20, no. 3, pp. 289-298, May 2011.

[7] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, B. Litt, “Epileptic Seizure Prediction Using Hybrid Feature Selection Over Multiple Intracranial EEG Electrode Contacts: A Report of Four Patients,” *IEEE Trans Biomed Eng*, vol. 50, no. 5, pp. 603-615, May 2003.

[8] N. Pop-Jordanova, and J. Pop-Jordanova, “Spectrum-weighted EEG frequency ("brain-rate") as a quantitative indicator of mental arousal,” *Prilozi*, vol. 26, no. 2, pp. 35-42, Dec 2005.

[9] P. C. Petrantonakis, and L. J. Hadjileontiadis, “Emotion Recognition From EEG Using Higher Order Crossings,” *IEEE Trans Inf Technol Biomed*, vol. 14, no. 2, pp. 186-197, Mar 2010.

[10] Q. Hu, Z. Xie, and D. Yu, “Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation,” *Pattern Recognition*, vol. 40, no. 12, pp. 3509-3521, Dec 2007

[11] S. Theodoridis, and K. Koutroumbas, *Pattern Recognition.* 3rd ed, San Diego, CA: Academic Press, 2006, pp. 235-236.