Expert Systems with Applications xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Classification of emotions induced by music videos and correlation with participants' rating

Daimi Syed Naser *, Goutam Saha

Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, Kharagpur 721 302, India

ARTICLE INFO

Keywords:
Dual-Tree Complex Wavelet Packet Transform (DT-CWPT)
Electroencephalogram (EEG)
Singular vector decomposition (SVD)
QR factorization with column pivoting (QRcp)
F-Ratio
Support vector machine (SVM)
Human–computer interaction (HCI)

ABSTRACT

Emotional experience and preference play a vital role in selection of multimedia content for an individual. Brain electrical activity bears the emotional cues needed for emotion detection, but very modest research has been done to extract those cues. This paper presents a novel machine learning approach using Dual-Tree Complex Wavelet Packet Transform (DT-CWPT) time–frequency features from electroencephalogram (EEG) to detect emotions together with an analysis of brain activity in different emotional states. Firstly, DT-CWPT is used to extract time–frequency emotional features. Then non-redundant and most discriminating emotional features are selected through singular value decomposition (SVD), QR factorization with column pivoting (QRcp) and F-Ratio based feature selection (FS) method. The reduced emotional feature set is used to classify emotion using support vector machine (SVM) and validated by leave-one-out cross-validation scheme. Results confirm the robustness and consistency in classification of emotions from EEG signals and significant correlation between participants' self assessed ratings with emotional features. It also gives an analysis of activities in brain region during different emotional states.

1. Introduction

Emotion plays a vital role in our daily life as it influences our intelligence, behavior and social communication. Knowledge of human emotions and its effects are crucial for the development in the field of affective computing that integrates emotions into human–computer interaction (HCI) (Picard). HCI needs to have emotional intelligence similar to human–human interaction. For this, HCI needs information regarding the human emotional experience and relation between emotional experience and the affective expression. Another area which can be benefit from this is multimedia content delivery (Koelstra et al., 2012; Lisetti & Nasoz, 2004; Soleymani, Lichtenauer, Pun, & Pantic, 2012). Multimedia items, such as songs and movies are liked by viewers' according to their emotional experience associated with it. And a viewer will search multimedia in order to fulfill the emotional need. However, now a day's multimedia items are tagged manually by meaningful word/words that describes their type, theme, factual contents or their affective features. An automatic implicit affective tagging would quickly tag media. A newly emerging area which can also benefit is Neuromarketing, by knowing the costumer's emotional choices and affective preferences of a product will helps to optimize the product (Khushaba et al., 2012, 2013). In above all applications, emotion recognition is the necessary and the most crucial step.

Emotion recognition techniques have been developed based on facial expression (Pantic & Rothkrantz, 2000), speech (Busso et al., 2004), and physiological signals like heart rate and galvanic skin response (GSR) (Lisetti & Nasoz, 2004; Picard, Vyzas, & Healey, 2001). Signals from central nerves system (CNS) (Electroencephalogram (EEG), Magnetoencephalogram (MEG), Position Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI)) are more reliable as compared to facial expression, speech and peripheral physiological signals from autonomous nervous system (ANS) (Petrantonakis & Hadjileontiadis, 2010a). For example, smile may not always represent positivity or agreement, they may even represent negativity or disagreement (Kappas, 2010). Moreover, the patient suffering from autism, emotion is not conveyed by facial expression. Similarly in speech, one may yell when angry, while one may speak politely even in anger. Whereas, ANS and CNS signals are usually not consciously controlled and they provide undisguised information of current emotional state (Tao & Tan). But, ANS signals like GSR shows similar properties as of emotion-driven e.g. inspiration from physical activity. EEG seems to be best suited among the CNS signals as it has better time resolution.

* Corresponding author. Tel.: +91 9163643888; fax: +91 3222255303.
E-mail addresses: sndaimi123@gmail.com (D.S. Naser), gsaha@ece.iitkgp.ernet.in (G. Saha).

http://dx.doi.org/10.1016/j.eswa.2014.03.050
0957-4174/

Notably, less research has been carried out on EEG signal based emotion recognition as compared to emotion recognition from facial, speech and peripheral physiological signals. Over the years a significant research is carried out on analysing EEG signals for emotion recognition where visual (Chanel, Kierkels, Soleymani, & Pun, 2009; Petrantonakis & Hadjileontiadis, 2010b), audio (Hadjidimitriou & Hadjileontiadis, 2012; Kim & Andre, 2008), or audiovisual (Bos; Ko, Yang, & Sim, 2009) stimuli are used. Moreover, very few uses music videos as stimuli and achieved notable performance. Takahashi (2004) used support vector machine (SVM) to classify emotion into five classes: anger, fear, joy, relaxation and sadness. In this work, EEG (FP1 and FP2 positions according to 10–20 system) and peripheral physiological signals (pulse rate and GSR) are used separately and together. Accuracy of 41.68% is obtained from EEG signals only and 41.72% from fusion of both in which 12 participants participated. Soleymani et al. (2012) have recorded multi-model emotional database in which 27 participants participated. In this work, EEG spectral features, ANOVA test to select discriminative features and SVM classifier are used to classify emotion on valance-arousal scale into three classes: low, medium and high. They achieved maximum classification accuracy of 57%. In other work Koelstra et al. (2012) have developed multi-model Dataset for Emotion Analysis using Physiological Signals (DEAP) where 32 participants participated. They used similar spectral features of EEG, Fisher's linear discriminant for feature selection and naïve Bayes classifier and achieved reasonable accuracy in classifying emotions as low/high arousal, valance and liking.

Most of studies are restricted to investigate feature extraction methods for representing emotional information from brain activity. Emphasis is given to spectral power of theta, alpha, beta and gamma band. There are few other studies concerned with non-stationary nature of EEG and use of time and frequency information as features. Murugappan et al. (2008) classified four emotions (disgust, happiness, surprise and fear) from 64-channel EEG signal recorded for six participants, where movie clips are used as stimuli. Here, Lifting based Discrete Wavelet Transform (DWT) with Daubechies wavelet of order 4 and 8 were used for feature extraction and only subband related to alpha band is used for classification using Fuzzy C-Means clustering (FCM). Event related potentials (ERPs) are used to classify emotion into positive and negative valance induced by visual stimuli (Hidalgo-Muñoz et al., 2013). Where, time–frequency features from Morlet wavelet and Support Vector Machine-Recursive Feature Elimination (SVM-RFE) are used for detecting scalp spectral dynamics of interest.

In this paper, we propose to use Dual-Tree Complex Wavelet Packet Transform (DT-CWPT) based energy features for classification of emotions from EEG signal of DEAP (Koelstra et al., 2012) database. Discrete Wavelet Transform (DWT) and Discrete Wavelet Packet Transform (DWPT) are widely used for feature extraction of EEG signals. But, this techniques suffer with non shift-invariant and aliasing issues arising from non-ideal filter and down sampling (Kingsbury, 2001). In order to overcome these disadvantages of DWT, Dual-Tree Complex Wavelet Transform (DT-CWT) was introduced by Kingsbury (2001) and DT-CWPT for DWPT (Bayram & Selesnick, 2008). Along with it, a suboptimal SVD–QRcp and F-Ratio based FS framework to reduce dimension and complexity of the system that delivers superior performance is used. This FS is found to be useful in subset selection in an application domain related to speech processing, speaker identification (Chakroborty & Saha, 2010) and heart sound classification (Ari & Saha, 2007). In addition to this, we explored how DT-CWPT subbands and brain regions are affected by different emotional states. For this, we did correlation analysis between EEG an subjective ratings of arousal, valance, liking and dominance.

The rest of the paper is organized as follows. Section 2 describes the structure of DT-CWPT decomposition in brief. SVD–QRcp and F-Ratio based feature selection technique is explained in detail in Section 3. Section 4 gives a description of DEAP database which is used for this work. Experimental results of single trial classification from EEG and correlation between participants self-assessed ratings and emotional features are analyzed and discussed in Section 5 and 6 respectively, and followed by conclusion in Section 7.

2. Dual-Tree Complex Wavelet Packet Transform (DT-CWPT)

Dual-Tree Complex Wavelet Packet Transform can be constructed from Dual-tree Complex Wavelet Transform (DT-CWT) by repeatedly decomposing each subband of first and second filter-bank (FB) (real and imaginary tree) using low-pass/high-pass perfect-reconstruction (PR) FBs. Each subband of the DT-CWPT will be analytic if the response of PR FBs of each branch of the second wavelet packet FB is the discrete Hilbert transform of the corresponding branch of the first wavelet packet FB (Bayram & Selesnick, 2008). The requirement to satisfy this is given next.

If a given filter $g(n)$ is the discrete Hilbert transform of some filter, $h(n)$, that is

$$G(e^{j\omega}) = jsgn(\omega)H(e^{j\omega}) \quad \text{for } |\omega| < \pi \tag{1}$$

then when $g(n)$ is convolved with some sequence $p(n)$, we have

$$G(e^{j\omega})P(e^{j\omega}) = jsgn(\omega)H(e^{j\omega})P(e^{j\omega}) \quad \text{for } |\omega| < \pi \tag{2}$$

As shown by Eq. (2), if $h(n)$ and $g(n)$ is a discrete Hilbert transform pair, then $g(n) * p(n)$ and $h(n) * p(n)$ are also discrete Hilbert transform pair (Bayram & Selesnick, 2008). The kth high-pass subband of second FB of DT-CWT is discrete Hilbert transform of kth high-pass subband of first FB. For wavelet packet this high-pass subband is to be further decomposed by using two-channel PR FBs.

Let $f_0(n), f_1(n)$ and $f_0'(n), f_1'(n)$ be the PR FB of first and second wavelet packet FB of DT-CWPT. From (1) and (2) the necessary condition on this filters $f_i(n), f_i'(n)$ such that the resulting DT-CWPT to be analytical is

$$f_i(n) = f_i'(n) \tag{3}$$

Fig. 1 shows the first wavelet packet FB of a four-stage DT-CWPT. The second wavelet FB is obtained by replacing the first stage filters $h_i^{(1)}$ by $h_i^{(1)}(n-1)$ and by replacing $h_i(n)$ by $h_i'(n)$ for $i \in \{0, 1\}$. The filters denoted by F_i are unchanged in the second wavelet packet FB. The frequency responses of four-stage DT-CWPT are shown in Fig. 2 which are approximately analytic.

3. SVD–QRcp and F-Ratio based feature selection

3.1. Singular value decomposition (SVD)

SVD is an optimal orthogonal decomposition in least square sense. SVD of a $m \times n$ matrix $\mathbf{A}$, given by

$$\mathbf{A} = \mathbf{USV}^T \tag{4}$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ is unitary matrix whose m columns are called left-singular vectors, $\mathbf{V} \in \mathbb{R}^{n \times n}$ is unitary matrix whose n columns are called right-singular vectors and $\mathbf{S}$ is the diagonal matrix with decreasing singular values of $\mathbf{A}$.

3.2. QR Factorization with column pivoting (QRcp)

QR factorization is a decomposition of a matrix $\mathbf{B}$ into a product

$$\mathbf{B} = \mathbf{QR} \tag{5}$$

Where $\mathbf{Q}$ is orthogonal matrix and $\mathbf{R}$ an upper triangular matrix. QR factorization with column pivoting means rotating the columns of the matrix $\mathbf{B}$ in order to maximize euclidian norm in successive

Q1 Please cite this article in press as: Naser, D. S., & Saha, G. Classification of emotions induced by music videos and correlation with participants' rating. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.03.050

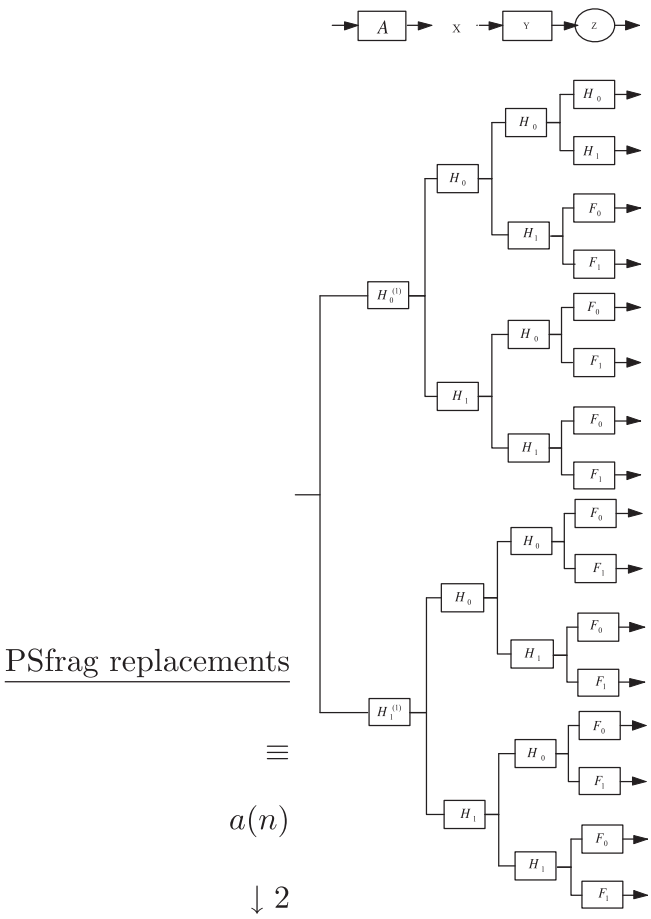

Fig. 1. First wavelet packet FB of a four-stage DT-CWPT.

directions, while QR factorization is performed on the matrix. The mechanism of the rotation of the columns is described in Kanjilal, Saha, and Koickal (1999). The sequence of the successive selection of the columns of **B** with maximum euclidian norm is recorded in a matrix called permutation matrix. The QR factorization with column pivoting is given as

$$\mathbf{B} = \mathbf{QRP}^T \tag{6}$$

Where **P** is the permutation matrix registering column swapping.

3.3. F-Ratio

F-Ratio is a statistical measure in the analysis of variance where multi-cluster data is available. If there are k number of clusters, and each cluster consists of n number of data points then

$$\text{F-Ratio} = \frac{\frac{1}{k}\sum_{j=1}^{k}(\mu_j - \bar{\mu})^2}{\frac{1}{k}\sum_{j=1}^{k}\frac{1}{n}\sum_{i=1}^{n}(a_{ij} - \mu_j)^2} \tag{7}$$

Where a_{ij} is the ith row of the jth class, μ_j is mean vector of the jth class and $\bar{\mu}$ is global mean vector of the data. F-Ratio will increase if the clusters move away from each other or the clusters shrink in their positions.

3.4. Feature selection

All features do not contain the same amount of information and few of them may be redundant and may not be useful for discrimination purpose. Here we use a combination of SVD, QRcp and F-Ratio based method (Ari & Saha, 2007; Chakroborty & Saha, 2010) for subset feature selection of input features. The entire process of feature selection is described as follows.

1. Let $m \times n$ be the dimension of input feature set **A** with $m \geqslant n$. The objective is to find an $m \times g$ subset $\mathbf{A}_1 (g < n)$ of **A**, which contains significant part of the information of **A**.
2. Perform singular value decomposition of **A** such that, $\mathbf{A} = \mathbf{USV}^T$.
3. Find g number of features for which the energy of singular values exhibit 99.5% of the total energy shown by n singular values.
4. QRcp factorization is performed on matrix **A** as in Chakroborty and Saha (2010) instead of that proposed by Ari and Saha (2007).
5. A combined ranking from QRcp and F-Ratio is obtained from

$$(\text{Combined Score})_i = \frac{(\text{QRcp Rank} + \text{F-Ratio Rank})}{2} \tag{8}$$

If there is any tie between two features, the feature having better QRcp ranking is given priority in selection.

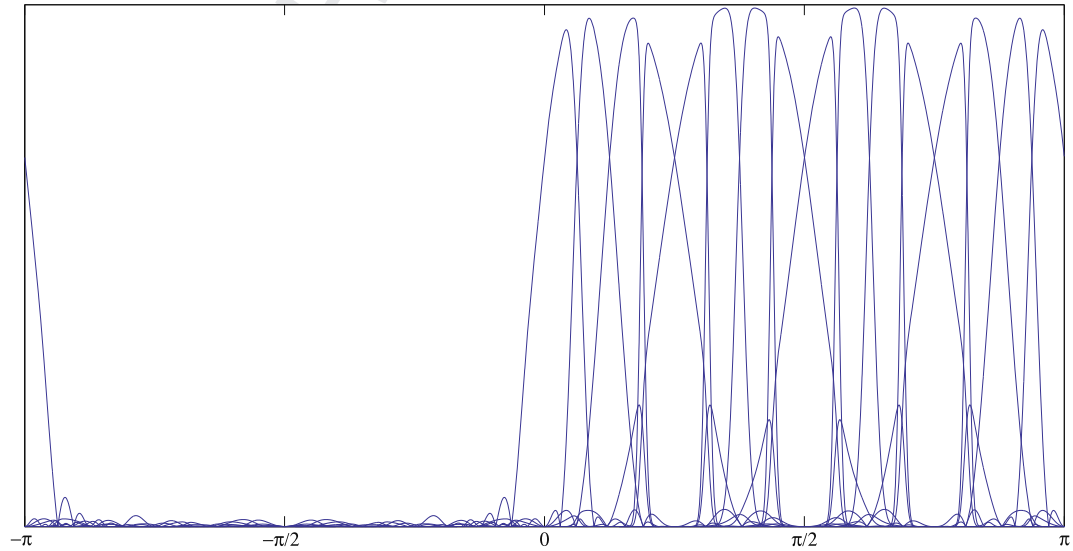

Fig. 2. Frequency domain analysis of DT-CWPT using Q-filters (Kingsbury, 2003) of length 14 and first stage $h_i^{(1)}$ and extension filters F_i as *db*14 wavelet filters.

Q1 Please cite this article in press as: Naser, D. S., & Saha, G. Classification of emotions induced by music videos and correlation with participants' rating. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.03.050

6. The reduced model discards $(n-g)$ number of low ranked features based on the combined QRcp and F-Ratio score. This selection gives sub-optimal result as optimal selection requires an exhaustive search which is prohibitively costly for large feature numbers. In this work, the number of features to start with is 552. Improvement in performance is sought by investigating the $(g-6)$ and $(g+6)$ neighbourhood to get best input features. Which gives superior result with relatively less computational cost (as search space is limited).

4. Database

DEAP dataset (Koelstra et al., 2012) is recorded over 32 participants aged between 19 and 37. Fifty percent of the participants are female and mean age is 26.9 years. Central nervous system, peripheral physiological signal and facial videos are recorded while participants watched 40 excerpts of selected music videos. Participants performed self-assessment of their levels of arousal, valance, liking and dominance at the end of each video. These videos are gathered using a novel stimuli selection method from relatively large set of music video clips followed by subjective test, to select the most appropriate test material. EEG signals are recorded from 32 active electrodes. Peripheral physiological signals are EMG, electrooculogram (EOG), blood volume pulse (BVP) using plethysmograph, skin temperature, and GSR. The face video collected is available for 22 of the 32 participants. Fig. 3 shows Electrode placement for 32-channel EEG recording using 10–20 system.

5. Single trial classification

In this section, we present the methodology and results of single trial classification of the emotions induced by music videos. Features from EEG signals of DEAP dataset (Koelstra et al., 2012) are used for classification. Participants' ratings during experiment are used as base to create four, two-class problems namely classification of low/high arousal, low/high valance, low/high liking and low/high dominance. The participants' ratings are separated into two classes (low and high) by selecting the threshold at 4.5 of 9-point self-assessment rating scale. The dataset is little unbalanced, since the percentage of videos belonging to high class per rating scale are: arousal 63.9%, valance 63.1%, liking 69.5% and dominance 66.8%.

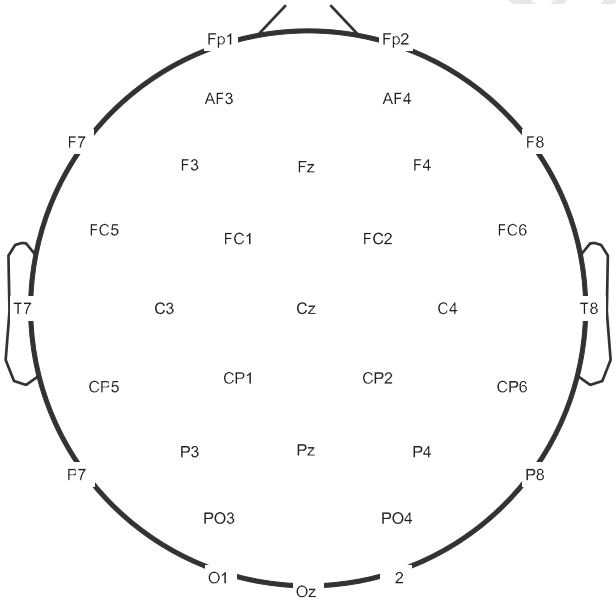

Fig. 3. Electrode placement for 32-channel EEG recording using 10–20 system. Channel labels: FP-pre-frontal, F-frontal, P-pareital, C-central, O-occipital, T-temporal, AF-anterior-frontal, FC-frontal-central, CP-central-parietal, PO-parietal-occipital, z-midline, odd numbers on the left, even numbers on the right of the subject.

The preprocessed EEG signals of DEAP dataset are common referenced, downsampled to 128 Hz, bandpass filtered with cutoff frequency 4–45 Hz using EEGLAB toolbox.[1] EOG artifacts are removed using blind source separation technique. Second harmonic of 50 Hz power line artifact is removed using notch filter. For feature extraction, EEG signals are decomposed using DT-CWPT. This at fourth level produces 16 subbands in each of real and imaginary trees. We select 12 subbands which fall in frequency range 0–48 Hz of each tree for feature extraction. Subbands of imaginary tree are Hilbert transform of subbands of real tree and magnitude of $((\text{Real-tree subband}) - j*(\text{Imaginary-tree subband}))$ is approximately analytical. Let this magnitude be H. We use logarithmic sum of magnitude square of H as the feature, which can be written as logarithm of energy of real tree subband plus energy of imaginary tree subband.

$$E_i = log\sum_{k=1}^{N}|H_i|^2 = log\sum_{k=1}^{N}|H_{ik}^R - jH_{ik}^I|^2 = log\sum_{k=1}^{N}(H_{ik}^R)^2 + (H_{ik}^I)^2 \quad (9)$$

where E_i, H_i^R and H_i^I for $i = 1, 2, 3, \ldots, 16$, are the logarithmic energy, real tree subband and imaginary tree subband respectively. And N is the number of wavelet coefficient of fourth level decomposed subband of DT-CWPT.

First, the features are extracted from EEG signals of each participant for each video. For feature extraction, each channel of 32-channel EEG is decomposed by DT-CWPT and 12 energy features are extracted as explained above. In addition, difference between energy features of all symmetrical pairs of electrodes on right and left cortical hemisphere are extracted to measure unevenness in brain activation in reaction to emotional stimuli. Thus, the total number of features per trial (video) is 552 (12-subband $*$ (32-electrodes + 14-electrode pair)). Then, feature selection is performed, as explained in Section 3,, to select a smaller but sub-optimal feature subset from 552 features. We use SVM classifier which is robust to class-imbalance situation (Nathalie & Shaju, 2002) with radial basis kernel. The value of gamma of radial basis function is set to default which is $1/num\_features$. F1-score and accuracy is used to evaluate performance of emotion classification in a leave-one-out cross validation scheme. F1-score is a measure which takes the class-balance into account. At each step of the cross validation, one video is used as the test-set and the rest are used as training-set.

Table 1 shows the average accuracies over participants and F1-score (average F1-score for both classes). The results of DT-CWPT features are compared with power spectral features of theta, alpha, slow alpha, beta and gamma bands (Koelstra et al., 2012). Fisher's linear discriminant is used to select discriminating features from the power spectral features. The threshold to select maximally discriminating features is empirically determined as 0.006. After feature selection from the power spectral features naive Bayes classifier is used for classification. Let A be the set of 32 channels and B be the set of 14 pairs of channels used for feature extraction. Table 2 shows EEG channel from set-A and set-B along with wavelet subband of the features selected from feature extraction method. The energy difference of symmetrical electrode pairs on the right and left hemisphere for example $AF4$ and $AF3$ is written as $AF4$–$AF3$. Fig. 4 shows the ROC plot of the classification result of DT-CWPT features.

[1] http://sccn.ucsd.edu/eeglab/

Table 1
Average accuracies (ACC) over participants, number of features (NoF) and F1-scores.

Features	Valance			Arousal			Like			Dominance		
	ACC	NoF	F1	ACC	NoF	F1	ACC	NoF	F1	ACC	NoF	F1
DT-CWPT (12-subbands) + (SVD, QRcp, F-Ratio) + SVM	0.653	19	0.550	0.669	14	0.570	0.712	16	0.509	0.691	7	0.552
Spectral Features + FLD + naive Bayes	0.616	4	0.492	0.571	15	0.559	0.641	38	0.542	0.532	30	0.525

Table 2
Features selected belongs to EEG channel and wavelet subband.

Valance		Arousal		Like		Dominance	
Subband	Channel	Subband	Channel	Subband	Channel	Subband	Channel
0–4 Hz	O2–O1	0–4 Hz	FC2–FC1	0–4 Hz	Oz	0–4 Hz	CP2
0–4 Hz	Oz	32–36 Hz	P4	0–4 Hz	FP2–FP1	4–8 Hz	FP2
44–48 Hz	O2–O1	0–4 Hz	P4–P3	0–4 Hz	CP2	12–16 Hz	F7
0–4 Hz	P4–P3	40–44 Hz	FC2	0–4 Hz	O1	0–4 Hz	C4–C3
44–48 Hz	P4–P3	40–44 Hz	T7	0–4 Hz	PO3	0–4 Hz	CP1
0–4 Hz	PO3	8–12 Hz	P4–P3	0–4 Hz	F8–F7	44–48 Hz	AF4–AF3
0–4 Hz	CP2	44–48 Hz	T8–T7	12–16 Hz	F7	0–4 Hz	AF4–AF3
4–8 Hz	P8	12–16 Hz	FC2–FC1	44–48 Hz	FP1		
0–4 Hz	FC5	0–4 Hz	T8–T7	8–12 Hz	PO3		
32–36 Hz	P4	0–4 Hz	Oz	4–8 Hz	P8		
0–4 Hz	CP1	44–48 Hz	FC2–FC1	8–12 Hz	FP2–FP1		
8–12 Hz	P4–P3	0–4 Hz	FC2	0–4 Hz	F7		
0–4 Hz	F3	12–16 Hz	C4–C3	8–12 Hz	Oz		
0–4 Hz	FC2–FC1	44–48 Hz	P4–P3	44–48 Hz	Oz		
8–12 Hz	O2–O1			44–48 Hz	Fz		
36–40 Hz	AF4			40–44 Hz	F7		
0–4 Hz	O1						
0–4 Hz	AF4–AF3						
4–8 Hz	FP2						

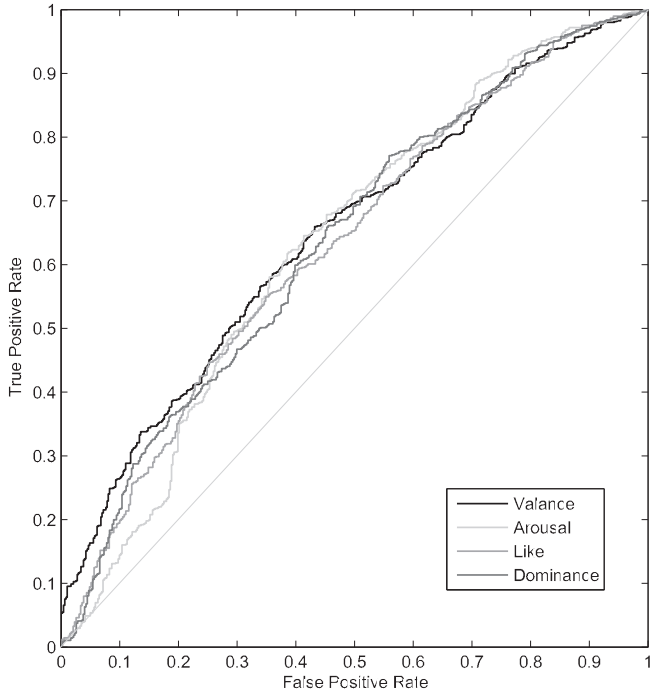

Fig. 4. ROC plot of classification results of DT-CWPT.

The topoplots of Fig. 5(a) and (b) show the position of selected channels for classifying the emotional valance. Out of 19 features selected for valance, 11 are from set-*A* and eight from set-*B*. Channels from set-*A* belong to right-frontal (FP2, AF4), left-frontal (F3, FC6), parietal (CP1, CP2, P4, P8), occipital (PO3, O1, O2) brain regions. These channels are from 0–4 Hz, 4–8 Hz, 32–36 Hz, and 36–40 Hz DT-CWPT subbands. Most of the subbands selected are from 0–4 Hz band. Set-*B* channels measure asymmetry between right and left hemisphere of frontal (AF4–AF3, FC2–FC1), parietal (C4–C3, P4–P3) and occipital (O2–O1) cortical areas and belong to 0–4 Hz, 8–12 Hz and 44–48 Hz subbands.

For emotional arousal classification, the selected channels are shown in topoplots of Fig. 5(c) and (d). The selected features are 14, out of which five are from set-*A* and remaining nine from set-*B*. The channels, left-temporal (T7), right fronto-central (FC2), right-parietal (P4) and occipital (Oz) are from set-*A*. This channel belong to 0–4 Hz, 32–36 Hz and 40–44 Hz subband. However, fronto-central (FC1–FC2), temporal (T7–T8) and parietal (P3–P4, C3–C4) channels are from set-B. These channels measure asymmetry in 0–4 Hz, 8–12 Hz, 12–16 Hz and 44–48 Hz subbands.

The selected channels for 'liking' classification are shown in topoplots of Fig. 5(e) and (f). Features selected mostly are from set-*A* and only three from set-B out of 16 features. Channels from cortical brain regions, left-frontal (Fp1, F7), frontal mid-line (Fz), right-parietal (CP2, P8), and occipital (O1, Oz, PO3) are from set-*A*. These features cover different subbands: 0–4 Hz, 4–8 Hz, 8–12 Hz, 12–16 Hz, 44–48 Hz, and 40–44 Hz. Asymmetry is measured only in channels of frontal (Fp2–Fp1, F8–F7) brain region and these belong to 0–4 Hz, 8–12 Hz and 44–48 Hz subbands. Fig. 5(g) and (h) shows the topoplot of dominance classification. Dominance classification uses less number of features as compared to all above classifications. The features from set-*A* are from right-frontal (Fp2), left-frontal (F7) and parietal (CP1, CP2) brain cortical regions. And belongs to 0–4 Hz, 4–8 Hz and 12–16 Hz subbands. Channels from frontal (AF4–AF3) and parietal (C4–C3) measure the dominance asymmetry which belong to 0–4 Hz and 44–48 Hz subbands.

It is observed that different electrodes of different brain regions are selected by feature selection method for valance, arousal, liking and dominance classification. Here, most of the electrodes selected

Q1 Please cite this article in press as: Naser, D. S., & Saha, G. Classification of emotions induced by music videos and correlation with participants' rating. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.03.050

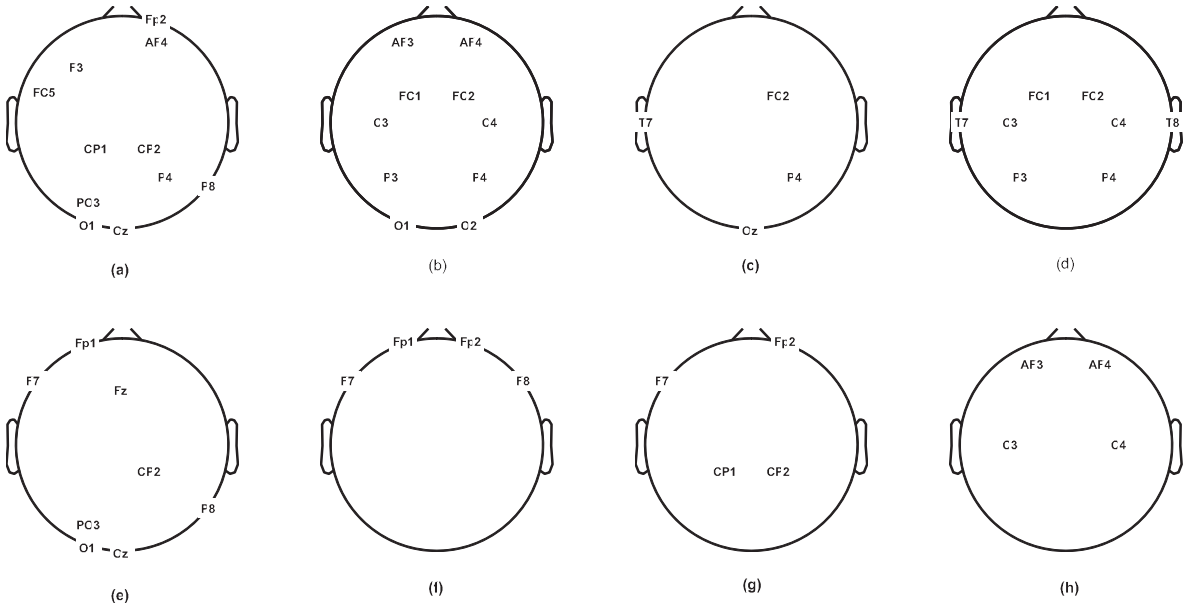

Fig. 5. (a), (c), (e), (h) channels selected from set-*A* and (b), (d), (f), (g) from set-*B* for valance, arousal, liking and dominance.

are from frontal, temporal and parietal lobes which are involved in emotion-related sensory activities reported in the literature on neurophysiological affective response. Some of the occipital lobes are also selected, which is basically involved in visual processing, it may be due to use of music video as stimuli.

6. Correlates of brain activity and ratings

In this section, correlation of EEG and corresponding subjective rating of arousal, valance, liking and dominance is analysed to investigate changes in brain regions due to different emotional states. In other words we examine how DT-CWPT subbands and brain regions are affected by different emotional states. The same preprocessed EEG signal of Section 5 is used for correlation analysis. The entire 60 s long signal is not used for this analysis but a prominent length is selected in order to remove bias of initial emotional state of participants. Now to find the prominent length the 60 s EEG signal of each participant is segmented into 25 s segment with 23 s overlap between segments (last segment is of 26 s duration). Accuracy of each segment is computed as in Section 5 and plotted as shown in Fig. 6. For correlation analysis, we select last 46 s signal where emotions are free from initial emotional state. Because we observe from Fig. 6 that for all (valance, arousal, liking and dominance) classification accuracy is stable from segment number nine.

For correlation statistics, the Spearman correlated coefficients between the energy and the subjective ratings are computed. Then, *p*-value is evaluated for left-(positive) and right-tailed (negative) correlation test. This is done for each participant separately and, assuming independance. The 32-resulting *p*-values per correlation direction (positive/negative), subband and electrode are combined to one *p*-value via Fisher's method. This technique combines a set of *p*-values from independent studies testing the same null hypothesis to collectively verify the null hypothesis. The null hypothesis in this case is the hypothesis that the logarithmic energy is uncorrelated with the subjective ratings of valance, arousal, like and dominance for each electrode and each participant.

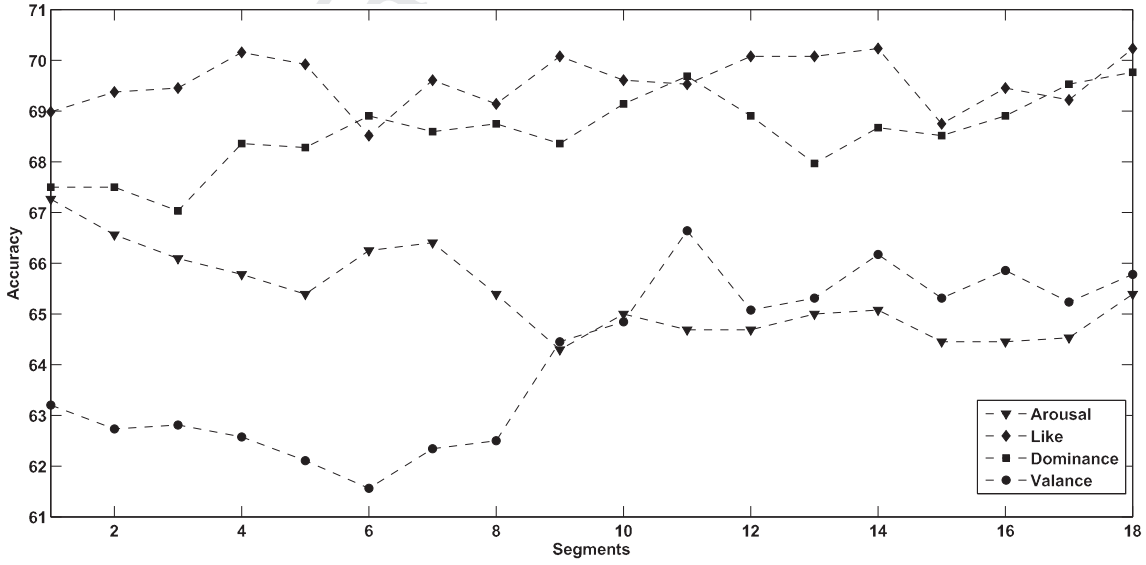

Fig. 6. Accuracies of 25 s segment with 23 s overlap.

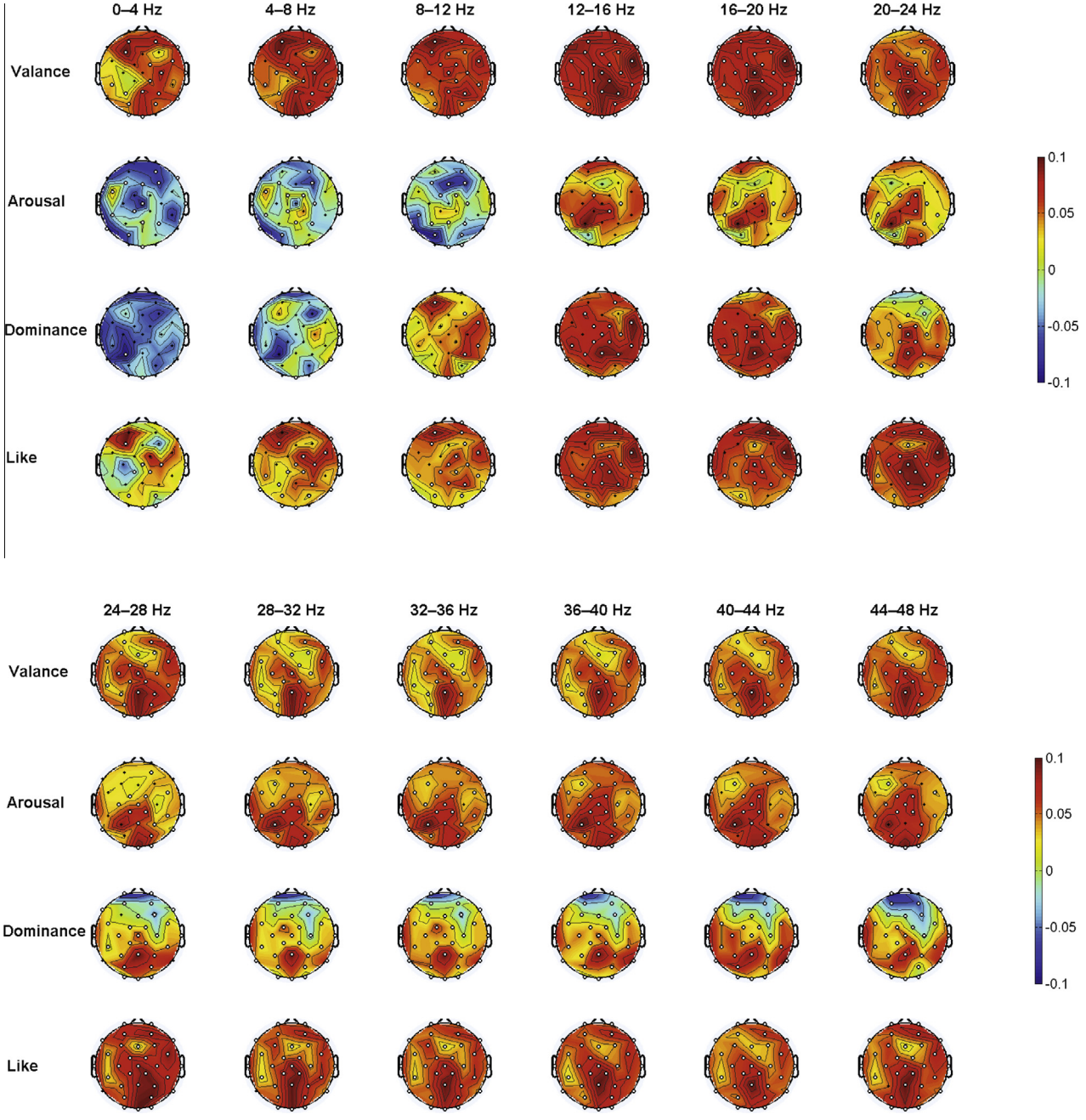

Fig. 7. The mean correlation (over all participants) of the valance, arousal, dominance and like, rating with the energy features of DT-CWPT subbands. The highlighted sensors correlate significantly ($p < 0.05$) with the ratings.

Hence, if the null hypothesis is rejected, there must be significant correlations between the logarithmic energy and the corresponding rating with significant p-value i.e, $p < 0.05$.

Fig. 7 shows the (average) correlations over participants. Significantly correlating electrodes ($p < 0.05$) are highlighted. Only these significant effects are sorted in ascending order and first three effects in each wavelet subband and emotional state are presented in Table 3. Arousal shows negative correlation in 0–4 Hz, 4–8 Hz and 8–12 Hz subbands. For higher arousal, energy in 0–4 Hz, 4–8 Hz and 8–12 Hz decreases. The 8–12 Hz subband, which is related to alpha band, a decrease in power with higher arousal is reported in Koelstra et al. (2012) and the reverse relation in Barry, Clarke, Johnstone, Magee, and Rushby (2007), this also follow here with respect to energy and arousal. Negative correlation is observed between energy and arousal as well as between energy and dominance in bands 0–4 Hz and 4–8 Hz. Arousal and liking follow opposite correlation patterns especially in 4–8 Hz and 8–12 Hz subband. This relates to he fact that participants shows preference for music videos that elicit low energy values for arousal (Kroupi, Yazdani, & Ebrahimi, 2011). There is not much change in the correlation patterns in the subbands related to gamma band i.e., subbands from 28–32 Hz to 44–48 Hz for any: valance, arousal, dominance and liking.

Valance shows the strongest correlations with emotional features in almost all DT-CWPT subbands analysed. An increase in valance leads to an increase in energy. This is similar to the finding in Koelstra et al. (2012). In subband 12–16 Hz related to low-beta band, an increase in occipital and temporal energy is associated

Q1 Please cite this article in press as: Naser, D. S., & Saha, G. Classification of emotions induced by music videos and correlation with participants' rating. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.03.050

Table 3
The first three significant electrode in each wavelet subband are highlighted. Also shows ($\bar{R}$) mean of subject-wise correlation, (R^-) the most negative and (R^+) the most positive.

	Valance				Arousal				Like				Dominance			
	Ch.	$\bar{R}$	R^-	R^+	Ch.	$\bar{R}$	R^-	R^+	Ch.	$\bar{R}$	R^-	R^+	Ch.	$\bar{R}$	R^-	R^+
0–4 Hz	T8	0.04	−0.45	0.53	CP5*	−0.01	−0.47	0.47	T8*	0.00	−0.38	0.51	Oz	−0.00	−0.4	0.5
	P3	0.01	−0.41	0.54	CP1*	−0.22	−0.46	0.33	CP2	0.01	−0.41	0.42				
	FC2	0.04	−0.44	0.51	OZ*	−0.00	−0.48	0.38	F3	0.04	−0.36	0.40				
4–8 Hz	T8**	0.07	−0.46	0.55	CP5*	−0.00	−0.47	0.46	T8*	0.03	−0.30	0.52	T8	0.00	−0.43	0.54
	Oz**	0.11	−0.38	0.52	Oz*	−0.00	−0.37	0.42	CP2	0.02	−0.39	0.42	Oz	−0.00	−0.36	0.56
	P3**	0.04	−0.43	0.49	C4*	−0.01	−0.64	0.38	C4	0.05	−0.38	0.45	P8	0.01	−0.37	0.51
8–12 Hz	P8**	0.11	−0.49	0.57	PO3**	−0.06	−0.59	0.48	PO3**	0.02	−0.39	0.50	P4*	0.06	−0.35	0.48
	FC2**	0.11	−0.46	0.51	P7*	−0.06	−0.60	0.36	O1**	0.02	−0.37	0.51	P8*	0.03	−0.35	0.61
	Oz**	0.15	−0.27	0.58	Oz*	−0.01	−0.49	0.48	CP2*	0.04	−0.37	0.48	FC6*	0.05	−0.45	0.41
12–16 Hz	Pz**	0.20	−0.35	0.68	FC1*	0.05	−0.38	0.44	CP2**	0.11	−0.31	0.60	AF4*	0.07	−0.38	0.66
	CP2**	0.19	−0.33	0.56	Pz	0.00	−0.38	0.56	FC2**	0.08	−0.36	0.59	FC6*	0.11	−0.33	0.50
	FC6**	0.21	−0.26	0.70	CP2	0.07	−0.36	0.59	AF4*	0.10	−0.31	0.57	CP2*	0.09	−0.51	0.60
16–20 Hz	FC6**	0.19	−0.26	0.76	FC1*	0.02	−0.59	0.48	AF4**	0.10	−0.45	0.66	AF4**	0.06	−0.47	0.68
	AF4**	0.16	−0.34	0.74	CP5	0.02	−0.50	0.47	O1**	0.05	−0.47	0.68	T8**	0.07	−0.39	0.62
	Pz**	0.19	−0.30	0.70	FC2	0.04	−0.57	0.41	FC6**	0.10	−0.41	0.68	CP1*	0.07	−0.44	0.59
20–24 Hz	FC2**	0.10	−0.44	0.66	FC1*	0.03	−0.43	0.42	PO3**	0.03	−0.53	0.67	FC2**	0.05	−0.43	0.58
	FC6**	0.13	−0.41	0.74	CP6	0.01	−0.48	0.50	CP2**	0.08	−0.48	0.69	AF4*	0.00	−0.44	0.42
	P4**	0.13	−0.46	0.62	PO3	−0.02	−0.49	0.38	P4**	0.08	−0.50	0.69	CP5*	0.04	−0.44	0.51
24–28 Hz	Oz**	0.11	−0.48	0.64	PO4*	0.07	−0.46	0.47	P4**	0.09	−0.47	0.60	AF4*	−0.01	−0.50	0.41
	AF4**	0.08	−0.48	0.71	Oz*	0.11	−0.48	0.47	PO3**	0.05	−0.50	0.60	FC2*	0.01	−0.48	0.53
	Pz**	0.13	−0.44	0.70	CP1*	0.07	−0.45	0.45	O2**	0.05	−0.44	0.53	Pz*	0.06	−0.41	0.53
28–32 Hz	Oz**	0.12	−0.49	0.64	CP6*	0.08	−0.51	0.54	O2**	0.06	−0.47	0.51	Pz**	0.07	−0.45	0.54
	Pz**	0.13	−0.49	0.65	PO4*	0.10	−0.49	0.49	AF4**	0.08	−0.47	0.58	P4*	0.02	−0.51	0.50
	AF4**	0.02	−0.53	0.66	CP2*	0.10	−0.47	0.46	AF3**	0.05	−0.51	0.58	FC2*	0.01	−0.47	0.51
32–36 Hz	Pz**	0.15	−0.53	0.66	CP6**	0.09	−0.52	0.52	AF4**	0.08	−0.48	0.59	Pz**	0.07	−0.45	0.57
	Oz**	0.11	−0.53	0.65	FP2**	0.09	−0.53	0.51	O2**	0.07	−0.48	0.50	P4*	0.02	−0.51	0.49
	AF4**	0.08	−0.55	0.70	PO4**	0.12	−0.52	0.45	Pz**	0.12	−0.48	0.59	FC6*	0.02	−0.46	0.47
36–40 Hz	Pz**	0.16	−0.55	0.65	FC1**	0.08	−0.56	0.54	P4**	0.08	−0.48	0.60	P4**	0.03	−0.54	0.61
	Oz**	0.11	−0.58	0.63	CP1**	0.10	−0.55	0.44	O2**	0.08	−0.44	0.50	Pz**	0.06	−0.55	0.59
	FC6**	0.11	−0.58	0.60	CP6**	0.06	−0.55	0.52	AF4**	0.08	−0.47	0.55	AF4**	−0.01	−0.54	0.40
40–44 Hz	Pz**	0.17	−0.55	0.66	CP6*	0.06	−0.53	0.52	Pz**	0.13	−0.50	0.59	AF4**	−0.01	−0.55	0.42
	Oz**	0.11	−0.54	0.68	Oz*	0.11	−0.53	0.44	PO3**	0.06	−0.51	0.47	FC1**	0.02	−0.55	0.48
	FC6**	0.11	−0.55	0.72	FC1*	0.06	−0.54	0.49	AF4**	0.09	−0.51	0.57	Pz**	0.05	−0.55	0.58
44–48 Hz	Pz**	0.14	−0.45	0.70	CP6*	0.05	−0.50	0.55	Pz**	0.09	−0.47	0.61	T8**	0.03	−0.50	0.46
	Oz**	0.11	−0.49	0.67	AF4*	0.06	−0.49	0.36	PO3**	0.06	−0.50	0.47	Pz**	0.04	−0.47	0.51
	FC6**	0.09	−0.50	0.72	O1*	0.09	−0.45	0.43	AF4**	0.08	−0.49	0.58	FC6**	0.00	−0.52	0.49

* $p < 0.01$.
** $p < 0.0001$.

with positive emotional self induction and external simulation (Koelstra et al., 2012). It is observed that a highly significant increase in parietal and occipital energy in subbands related to high-beta and gamma.

The liking correlates are found in almost all analysed DT-CWPT subbands. For 0–4 Hz, 4–8 Hz and 8–12 Hz subbands, increase of energy over left frontal-central and anterior-frontal cortex is observed. Increase in parietal and occipital energy in subbands related to high-beta and gamma bands is similar to those observed for valance. There seems to exist much similarity between correlations of valance and liking, which might be the result of high inter-correlation of scales. Dominance shows negative correlation in low frequency subbands (0–4 Hz and 4–8 Hz). Energy in these bands increases with low dominance. Subbands related to alpha and beta shows positive correlation. Where as, there is negative correlation in pre-frontal cortex and positive correlation in occipital and right parietal cortex in all subbands related to high-beta and gamma.

7. Conclusion

In this work, we have presented a novel method for emotion recognition induced by music videos. For this we propose to use DT-CWPT time–frequency features from brain electrical activity. In addition to that, SVD–QRcp and F-Ratio based feature selection method is employed for eliminating weak and redundant features. We observe that the number of features selected from this technique are less than 4% of total number of features (552). With these features the average classification rates are 65.3% for valance, 66.9% for arousal, 71.2% for like and 69.1% for dominance. This show that emotional features extracted using DT-CWPT represent user's emotional state effectively. While analysing the selected features we notice. First, we found that different electrodes of different brain regions are selected for valance, arousal, liking and dominance. Second, energy features representing asymmetry between the right and left hemisphere of brain are prominent in valance and arousal classification as compared to liking and dominance. Third, temporal lobe is involved in arousal classification only. Fourth, different subbands play significant role in different brain regions. In addition to this, we explored how DT-CWPT subbands and brain regions are activated by different emotional states. It was observed that, emotional features over several brain regions are correlated to preferences of the participants. Valance and liking shows strongest correlation in all analysed subbands. Whereas, there is significant correlation in parietal and occipital brain regions related to high-beta and gamma subbands for all: valance, arousal, liking and dominance. An opposite correlation patterns in 4–8 Hz and 8–12 Hz subbands is also noticed for arousal and liking.

Q1 Please cite this article in press as: Naser, D. S., & Saha, G. Classification of emotions induced by music videos and correlation with participants' rating. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.03.050

In future work, we aim to investigate phase information of DT-CWPT subbands in emotion classification and its affects on different brain regions. Further, we like to find optimal electrode positions for emotion recognition.

References

Ari, S., & Saha, G. (2007). In search of an SVD and QRcp based optimization technique of ANN for automatic classification of abnormal heart sounds. *International Journal of Biomedical Sciences, 2*(1), 1013–1016.

Barry, R., Clarke, A., Johnstone, S., Magee, C., & Rushby, J. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology, 118*(12), 2765–2773.

Bayram, I., & Selesnick, I. W. (2008). On the dual-tree complex wavelet packet and M-band transforms. *IEEE Transactions on Signal Processing, 56*(6), 2298–2310.

Bos, D. EEG-based emotion recognition: The influence of visual and auditory stimuli. <http://hmi.ewi.utwente.nl/verslagen/capita-selecta/>.

Busso, C., Deng, Z., Yildirim, S., Buut, M., Lee, C., Kazemzadeh, A., Lee, S., Neumann, U., & Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proc. sixth int'l conf. multimodal interfaces* (pp. 205–211).

Chakroborty, S., & Saha, G. (2010). Feature selection using singular value decomposition and QR factorization with column pivoting for text-independent speaker identification. *Speech Communication, 52*(9), 693–709.

Chanel, G., Kierkels, J., Soleymani, M., & Pun, T. (2009). Short-term emotion assessment in a recall paradigm. *International Journal of Human–Computer Studies, 67*(8), 607–627.

Hadjidimitriou, S. K., & Hadjileontiadis, L. J. (2012). Toward an EEG-based recognition of music liking using time–frequency analysis. *IEEE Transactions on Biomedical Engineering, 59*(12), 3498–3510.

Hidalgo-Muñoz, A., López, M., Santos, I., Pereira, A., Vázquez-Marrufo, M., Galvao-Carmona, A., & Tomé, A. (2013). Application of SVM-RFE on EEG signals for detecting the most relevant scalp regions linked to affective valence processing. *Expert Systems with Applications, 40*(6), 2102–2108.

Kanjilal, P. P., Saha, G., & Koickal, T. J. (1999). On robust nonlinear modeling of a complex process with large number of inputs using m-QRcp factorization and C_p statistic. *IEEE Transactions on Systems, Man, and Cybernetics Part B, 29*(1), 1–12.

Kappas, A. (2010). Smile when you read this, whether you like it or not: Conceptual challenges to affect detection. *IEEE Transactions on Affective Computing, 1*(1), 1672–1687.

Khushaba, R. N., Greenacre, L., Kodagoda, S., Louviere, J., Burke, S., & Dissanayake, G. (2012). Choice modeling and the brain: A study on the Electroencephalogram (EEG) of preferences. *Expert Systems with Applications, 39*(16), 12378–12388.

Khushaba, R. N., Wise, C., Kodagoda, S., Louviere, J., Kahn, B. E., & Townsend, C. (2013). Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking. *Expert Systems with Applications, 40*(9), 3803–3812.

Kim, J., & Andre, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transaction on Pattern Analysis and Machine Intelligence, 30*(12), 2067–2083.

Kingsbury, N. G. (2001). Complex wavelets for shift invariant analysis and filtering of signals. *Journal of Applied and Computational Harmonic Analysis, 10*(3), 234–253.

Kingsbury, N. G. (2003). Design of Q-shift complex wavelets for image processing using frequency domain energy minimization. In *Proc. IEEE int. conf. image process (ICIP)* (pp. 1013–1016).

Koelstra, S., Muehl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing, 3*(1), 18–31.

Ko, K.-E., Yang, H.-C., & Sim, K.-B. (2009). Emotion recognition using EEG signals with relative power values and bayesian network. *International Journal of Control, Automation and Systems, 7*(5), 865–870.

Kroupi, E., Yazdani, A., & Ebrahimi, T. (2011). EEG correlates of different emotional states elicited during watching music videos. In *Proc. 4th int. conf. affective comput. intell. interact.*, Memphis, TN (pp. 457–466).

Lisetti, C., & Nasoz, F. (2004). Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Advances in Signal Processing, 11*, 1672–1687.

Murugappan, M., Rizon, M., Nagarajan, R., Yaacob, S., Zunaidi, I., & Hazry, D. (2008). Lifting scheme for human emotion recognition using EEG. In *Proc. int. symp. inf. technol.* (pp. 1–7).

Nathalie, J., & Shaju, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6*(5), 429–449.

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1424–1445.

Petrantonakis, P., & Hadjileontiadis, L. (2010a). Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on Affective Computing, 1*(2), 1672–1687.

Petrantonakis, P., & Hadjileontiadis, L. (2010b). Emotion recognition from EEG using higher order crossings. *IEEE Transactions on Information Technology in Biomedicine, 14*(2), 1–6.

Picard, R. *Affective computing*. Cambridge, MA: MIT Press.

Picard, R., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*, 1175–1191.

Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A multimodal affective database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing, Special Issue on Naturalistic Affect Resources for System Building and Evaluation, 3*(1), 1–6.

Takahashi, K. (2004). Remarks on emotion recognition from bio-potential signals. In *Proc. of 2nd int. conf. autonomous robots agents* (pp. 186–191).

Tao, J., & Tan, T. *Affective information processing*, Springer.

Q1 Please cite this article in press as: Naser, D. S., & Saha, G. Classification of emotions induced by music videos and correlation with participants' rating. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.03.050