35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

EEG-based Recognition of Video-induced Emotions: Selecting Subject-independent Feature Set

Jukka Kortelainen, *Member, IEEE*, and Tapio Seppänen

Abstract— **Emotions are fundamental for everyday life affecting our communication, learning, perception, and decision making. Including emotions into the human-computer interaction (HCI) could be seen as a significant step forward offering a great potential for developing advanced future technologies. While the electrical activity of the brain is affected by emotions, offers electroencephalogram (EEG) an interesting channel to improve the HCI. In this paper, the selection of subject-independent feature set for EEG-based emotion recognition is studied. We investigate the effect of different feature sets in classifying person's arousal and valence while watching videos with emotional content. The classification performance is optimized by applying a sequential forward floating search algorithm for feature selection. The best classification rate (65.1% for arousal and 63.0% for valence) is obtained with a feature set containing power spectral features from the frequency band of 1-32 Hz. The proposed approach substantially improves the classification rate reported in the literature. In future, further analysis of the video-induced EEG changes including the topographical differences in the spectral features is needed.**

I. Introduction

EMOTIONS are fundamental for everyday life affecting our communication, learning, perception, and decision making. While playing a central part in the human-to-human communication, including emotions into the human-computer interaction (HCI) could be seen as a significant step forward offering a great potential for developing advanced future technologies. The task is, however, challenging as our scientific knowledge about the topic is still limited. Novel approaches for automatic recognition, processing, interpretation, and simulation of human emotions are therefore needed.

Humans express emotions via different channels, such as facial expressions, speech, and gestures. Numerous methods utilizing these modalities for automatic emotion recognition have been proposed (see for example [1]-[3]). In addition, physiological signals, such as heart rate variability (HRV), galvanic skin response (GSR), and electroencephalogram (EEG) are considered to contain essential information about the person's emotional state [4]-[6]. Consequently, technologies combining different modalities for automatic emotion recognition have recently gained considerable scientific interest. These include, for example, multimodal approaches using facial expressions with audio signals, different kinds of physiological signals, and facial expressions with physiological signals [7]-[9].

The electrical activity of the brain is affected by emotions. The frontal asymmetrical brain activity was shown to discriminate between positive and negative emotions more than 30 years ago [10] and after that an extensive amount of information about the neurobiological background of emotions have been published. In the EEG-based technological solutions, the emotions have conventionally been tracked by assessing the activity changes in the classical frequencies, i.e. delta (1-4 Hz), theta (5-8 Hz), alpha (9-12 Hz), beta (13-30), and gamma (>30 Hz) bands, as well as the activity differences between the hemispheres. The novel easy-to-attach and even wireless measurement systems have made EEG an interesting possibility for multimodal HCI. However, as the number of electrodes and derived parameters in these set-ups is usually high, extracting the relevant information from the huge amount of data is essential to improve the system's performance.

In this paper, the selection of subject-independent feature set for EEG-based emotion recognition is studied. We investigate the effect of different feature sets in classifying person's arousal and valence while watching videos with emotional content. The classification performance is optimized by applying a sequential forward floating search algorithm for feature selection. The structure of the paper is as follows: Section II describes the experimental protocol and used data as well as the feature extraction and the application of the feature selection algorithm. The results are given in Section III. In Section IV, the paper is concluded with some discussion about the achieved results compared to literature and future work.

II. Materials and Methods

A. Experimental Protocol and Data

The research in this paper uses the MAHNOB Database collected by Professor Pantic and the iBUG group at Imperial College London, and in part collected in collaboration with Professor Pun and his team of University of Geneva, in the scope of MAHNOB project financially supported by the European Research Council under the European Community's 7th Framework Programme (FP7/2007-2013) / ERC Starting Grant agreement No. 203143 [11].

Manuscript received February 4, 2013. This work was supported in part by grant 40297/11 from Tekes.

J. Kortelainen is with the Department of Computer Science and Engineering, BOX 4500, FIN-90014 University of Oulu, Finland (e-mail: jukka.kortelainen@ee.oulu.fi).

T. Seppänen is with the Department of Computer Science and Engineering, University of Oulu, Finland.

In the experimental protocol, 20 video clips were shown to 30 subjects (17 females, 13 males; 19-40 years old) with different cultural backgrounds. The video clips, whose duration was 34.9-117 s, were taken from commercially produced movies (14) or online resources (6). After seeing each clip, the subject picked a keyword for the emotion that best described the video's content. The keywords used were: sadness, joy/happiness, disgust, neutral, amusement, anger, fear, surprise, and anxiety. Each keyword was then mapped into one of three classes according to arousal and valence as presented in [12]. For arousal the classes were calm (sadness, disgust, neutral), medium arousal (joy/happiness, amusement), and excited/activated (anger, fear, surprise, anxiety). For valence the classes were unpleasant (sadness, disgust, anger, fear, anxiety), neutral valence (neutral, surprise), and pleasant (joy/happiness, amusement).

During the experiment, EEG was recorded using 32 active AgCl electrodes. The electrode locations, illustrated in Fig. 1, followed the international 10/20 system. The recording was performed first with a sampling frequency of 1024 Hz and the signals were then downsampled to 256 Hz. Due to unfinished data collection and technical problems, only 547 of the 600 data recordings (20 clips × 30 subjects) were obtained. In addition, six recordings were excluded from the analysis due to poor signal quality.

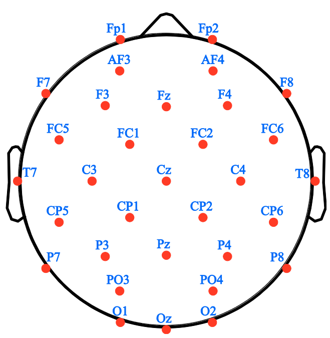

Fig. 1. The electrode locations used in the EEG recording. Figure created using EEGLAB [13]

B. EEG Feature Extraction

The processing of EEG signals as well as the data analysis presented in this paper was carried out using the Matlab technical computing language (The MathWorks Inc., Natick, MA).

The power spectral density (PSD) of each EEG recording was calculated using Welch's method. Window length was set to 5 s and overlap to 4 s. The PSDs were determined only for the signals recorded while the subjects were looking at the video excluding the self-evaluation parts. Three different feature sets were then created. The first one (FS1) contained powers in the conventional frequency bands: delta (1-4 Hz), theta (5-8 Hz), alpha (9-12 Hz), beta (13-30 Hz), and gamma (31-49 Hz). The second feature set (FS2) contained the powers in all single frequencies of the band 1-32 Hz. In addition, the powers in all adjacent 2 Hz, 4 Hz, 8 Hz, and 16 Hz wide frequency bands were included as well as the total power in the 1-32 Hz band. The third feature set (FS3) was constructed similarly than FS2, but the frequency band used was 1-48 Hz. The features were calculated for all 32 channels recorded. In addition, the power differences were determined for the 14 electrode pairs located symmetrically over the left and right hemispheres. The total number of features in FS1, FS2, and FS3 were thus 230, 2898, and 4324, respectively. To reduce the inter-individual variation, the features were normalized as follows. For each subject, the feature values were separately mapped to the range [0, 1]. This was done by subtracting the minimum value of the feature from all the feature values and then dividing the values by the difference between the maximum and minimum feature values.

C. Feature Selection and Classification

To reduce the feature sets and improve the classification performance, a sequential forward floating search method was applied to the data [14]. The method is based on a sequential search of the best feature subset using dynamic inclusion and exclusion of features. The algorithm contains the following steps:

1) Feature inclusion. Each feature that is not included in the feature set is tested and the one that leads to the best performance is included.

2) Conditional feature exclusion. Each feature that is included in the feature set is tested and if there is a feature whose removal leads to better performance compared to the performance received with the reduced feature set earlier, the feature is excluded. If more than one feature fulfills this criterion, the one that results in the best performance is selected.

In the beginning, none of the features is included in the feature set. Step 2 is performed after step 1 and is repeated until the criterion is not fulfilled. After this, the algorithm goes back to step 1. The algorithm stops when all or a predetermined number of features are included. It has been shown, that the algorithm provides an optimal or close to optimal solution while being computationally effective compared to the exhaustive approach of trying all the feature combinations. The classification rate was used as the measure of performance. The classification was performed with a k nearest neighbors (KNN) leave-one-subject-out approach.

The original number of features in FS2 and FS3 was high and the application of the feature selection method would have demanded too much time. Therefore, a preselection of features was carried out for the whole dataset using one-way ANOVA test. The test was done for each feature in both sets with the class (arousal and valence) as the independent variable. Only the features for which the p value was lower than a predefined threshold (0.05, 0.1, 0.15, or 0.2) were included when the above described feature selection algorithm was applied to the data.

TABLE I
FEATURE SELECTION AND CLASSIFICATION RESULTS

Feature set	Preselection criterion	Arousal Features, total	used	k	Classification rate	Valence Features, total	used	k	Classification rate
FS1	-	230	56	5	59.7%	230	103	3	55.8%
FS2	$p < 0.05$	322	39	3	61.4%	281	72	3	58.6%
	$p < 0.10$	536	57	7	60.8%	496	92	13	60.1%
	$p < 0.15$	713	90	11	62.9%	682	81	3	61.7%
	$p < 0.20$	896	90	3	**65.1%**	866	181	3	**63.0%**
FS3	$p < 0.05$	477	63	5	60.6%	397	82	11	58.8%
	$p < 0.10$	787	98	11	62.5%	685	51	17	60.8%
	$p < 0.15$	1054	77	15	62.5%	934	43	17	60.6%

III. RESULTS

The effect of preselection on the number of features is given in Table I. As the number of features in FS1 was rather low, no preselection was carried out for it. For FS3, the preselection criterion $p < 0.2$ resulted in too big feature set for the feature selection algorithm and was thus excluded. The effect of preselection on the relative contribution of different frequencies in feature set FS2 is illustrated in Fig. 2. For arousal, the low frequencies (< 6 Hz) as well as higher alpha frequencies (11-12 Hz) were emphasized in the feature set after preselection. In addition, the contribution of higher frequencies (> 20 Hz) was rather high with arousal. For valence, the lower frequencies (< 12 Hz) were slightly emphasized in the feature set and the contribution of higher frequencies was smaller than with arousal.

Table I also presents the best classification results obtained with the three feature sets and different preselection criteria. The best classification rate for arousal was 65.1% and for valence 63.0%. They were both achieved with feature set FS2, preselection criterion $p < 0.20$, and $k = 3$. For FS3, the best results for arousal (62.5%, $k = 15$) and valence (60.6%, $k = 17$) were obtained with preselection criterion $p < 0.15$. The features of FS1 representing the conventional frequency bands resulted in substantially lower classification rate (59.7% for arousal and 55.8% for valence).

In Fig. 3, the results of the feature selection algorithm are presented for feature set FS2 as a function of number of features. As shown already in Table I, the best classification rate for arousal was obtained with 90 features. For valence, 181 features resulted in the best classification rate.

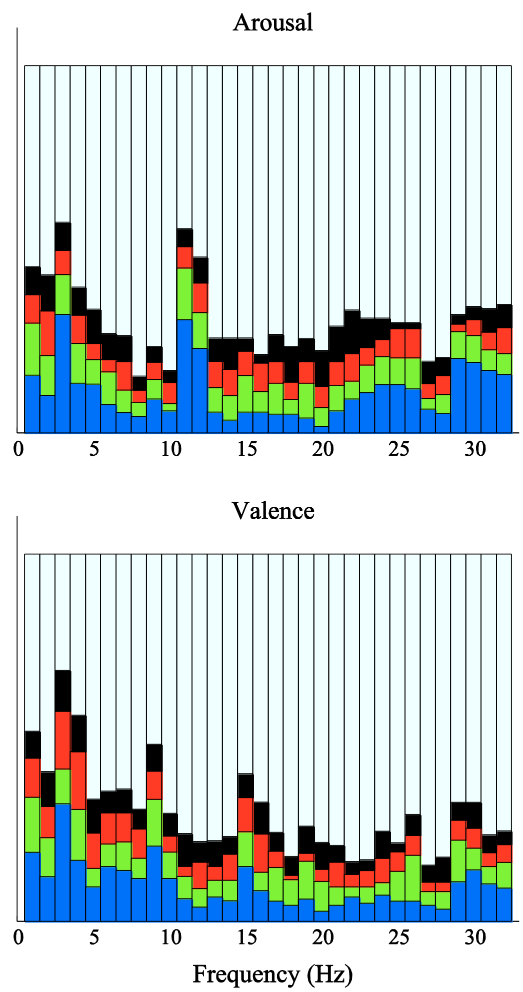

Fig. 2. The effect of preselection on the contribution of different frequencies in the feature set FS2 before the application of feature selection algorithm. White, black, red, green, and blue bars indicate the relative contribution of each frequency before preselection and after preselection using $p < 0.2$, $p < 0.15$, $p < 0.1$, and $p < 0.05$, respectively. The histograms are created so that each feature equally contributes to the area. The feature representing a single frequency increases the corresponding bin. The feature containing more than one frequency, for example a 2 Hz wide band, increases all the corresponding bins within the band. However, the bins are increased only half compared to the previous case.

IV. DISCUSSION

The proposed approach for EEG-based video-induced emotion classification substantially improves the results presented in the literature. When publishing the database used also in this study, Soleymani *et al.* reported a classification rate of 52.5% for arousal and 57.0% for

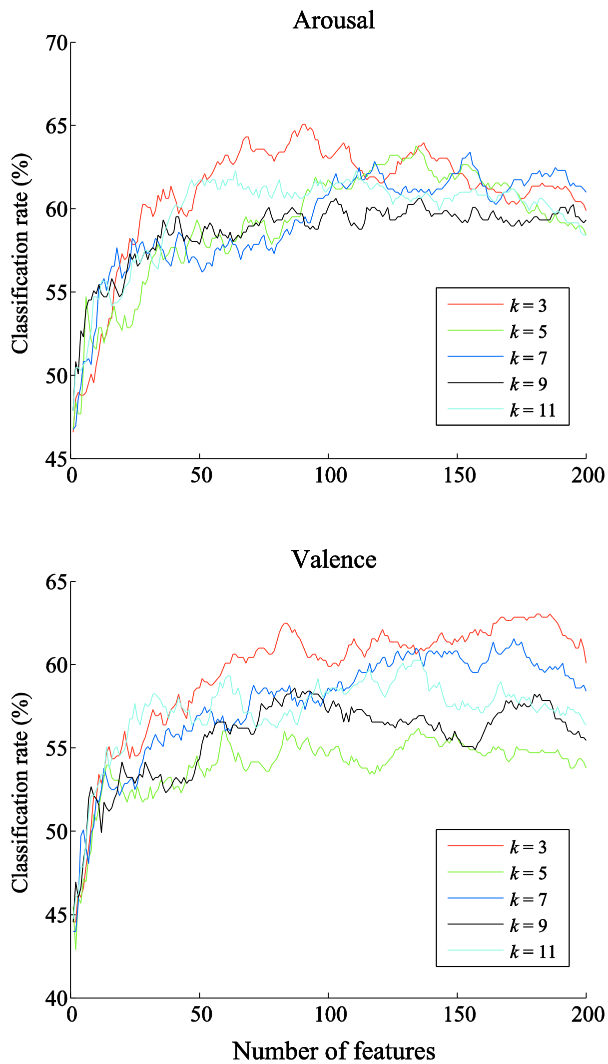

Fig. 3. The results of the feature selection algorithm when feature set FS2 was used with a preselection criterion $p < 0.2$. The classification rate is given as a function of number of features for five different k values.

valence when only EEG was used [11]. The results are comparable with this study as the approach for classification (leave-one-subject-out classification with keyword-based determination of arousal and valence for the video samples) was similar.

In future, further analysis of the EEG changes related to video-induced emotions is needed. This study illustrated the significance of the feature selection in the EEG-based classification of arousal and valence. However, reliable measurement of these parameters should be based on the neurophysiological phenomena related to the emotions induced. Finding out the phenomena requires a comprehensive topographical analysis of the EEG spectral changes.

V. Conclusions

In this paper, the selection of subject-independent feature set for EEG-based emotion recognition was studied. The effect of different feature sets in classifying person's arousal and valence while watching videos with emotional content was investigated. The classification performance was optimized by applying a sequential forward floating search algorithm for feature selection. The best classification rate (65.1% for arousal and 63.0% for valence) was obtained with a feature set containing power spectral features from the frequency band of 1-32 Hz. The proposed approach was shown to substantially improve the classification rate presented in the literature. In future, a topographical analysis of the essential spectral features for video-induced emotion classification is needed.

References

[1] A. Chakraborty, A. Konar, U. Chakraborty, and A. Chatterjee, "Emotion recognition from facial expressions and its control using fuzzy logic," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 39, pp. 726–743, 2009.

[2] E. Väyrynen, J. Kortelainen, and T. Seppänen, "Classifier-based learning of nonlinear feature manifold for visualization of emotional speech prosody," *IEEE Transactions on Affective Computing*, in press.

[3] D. Glowinski, A. Camurri, G. Volpe, N. Dael, and K. Scherer, "Technique for automatic emotion recognition by body gesture analysis," in *Proc. IEE CS Conf. Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1–6.

[4] G. Chanel, K. Ansari-Asl, and T. Pun, "Valence-arousal evaluation using physiological signals in an emotion recall paradigm," in *Proc. IEEE SMC*, 2007, pp. 2662–2667.

[5] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: first steps towards an automatic system," in *Affective Dialogue Systems, Tutorial and Research Workshop 2004*, 2004, pp. 36–48.

[6] Y. Lin, J. Chen, J. Duann, C. Lin, and T. Jung, "Generalizations of the subject-independent feature set for music-induced emotion recognition," in *Proc. 33rd Annu. Int. Conf. IEEE EMBS*, Boston, USA, 2011, pp. 6092–6095.

[7] H. Gunes, B. Schuller, M. Pantic, and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: a survey," in *Proc. IEEE International Conference in Automatic Face & Gesture Recognition*, Santa Barbara, CA, 2011, pp. 827–834.

[8] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: a review", in *Proc. IEEE 7th International Colloquium on Signal Processing and its Applications*, Penang, Malaysia, 2011, pp. 410–415.

[9] J. Kortelainen, S. Tiinanen, X. Huang, X. Li, S. Laukka, M. Pietikäinen, and T. Seppänen, "Multimodal emotion recognition by combining physiological signals and facial expressions: a preliminary study," in *Proc. 34th Annu. Int. Conf. IEEE EMBS*, San Diego, USA, 2012, pp. 5238–5241.

[10] R. Davidson and N. Fox, "Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants," *Science*, vol. 218, pp. 1235–1237, 1982.

[11] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multi-modal affective database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, pp. 42–55, 2011.

[12] J. Fontaine, K. Scherer, E. Roesch, and P. Ellsworth, "The world of emotions is not two-dimensional," *Psychol. Sci.*, vol. 18, pp. 1050–1057, 2007.

[13] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics," *J. Neurosci. Methods*, vol. 134, pp. 9-21, 2004.

[14] P. Pudil, H. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, pp. 1119–1125, 1993.