2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

ASSESSMENT OF NEW SPECTRAL FEATURES FOR EEG-BASED EMOTION RECOGNITION

Anne-Claire Conneau, Slim Essid

Institut Mines-Telecom/Telecom ParisTech, CNRS-LTCI, Paris, France

ABSTRACT

The choice of appropriate features for automatic emotion recognition based on electroencephalographic (EEG) signals remains to date an open research question. In this paper we explore a wide range of potentially useful features, including original ones, comparing them to previous proposals through a rigorous experimental evaluation, using a strict cross-validation protocol. In particular we assess the effectiveness of new spectral features–both in multi-channel and single-channel EEG setups–for the problem of discriminating positively and negatively excited emotions. The evaluation is conducted using the ENTERFACE'06 dataset allowing us to study the behaviour of the tested features across different subjects. Our results prove the usefulness of various new spectral features even in single-channel setups. We also observe that the optimal selection of features is highly subject-dependent. Finally combining different groups of features we find the valence recognition accuracy to be possibly as high as 78%.

***Index Terms*—** EEG, emotion recognition, valence, spectral features, common spatial patterns.

1. INTRODUCTION

While the usage of electroencephalographic (EEG) recording has been for long confined in the medical field, the recent years have seen a growing interest in EEG-based brain-computer interfaces (BCI) for general public applications. In particular EEG-recording has attracted the attention of researchers in the field of affective computing as part of the effort to perform human-behaviour analysis tasks, especially automatic emotion recognition. Compared to other modalities which have been considered in previous work on emotion recognition, such as speech, facial expressions, gestures or other physiological signals [1, 2, 3], EEG has the advantage of capturing information related to internal emotional states not necessarily resulting in any observable external manifestations (especially through the audio, visual or motion modalities).

Emotion recognition is usually approached as a classification problem where the choice of appropriate *features* is critical to ensure satisfactory recognition accuracy. As far as EEG-features are concerned, a consensus has not yet been reached as to a standard set of attributes that could guarantee a successful characterisation of a human-subject's emotions. It is nevertheless acknowledged in the field of neuroscience that a great deal of relevant information is conveyed by the spectral properties of the EEG signals, where different types of human activities result in different spectral patterns appearing in well-specified frequency bands.

The authors would like to thank Gaël Richard for his advice .
This research has been partially financed by the European Project VERVE n.288914 of the European Community Seventh Framework Program (FP7/2007-2013).

In this work we explore a wide range of *temporal*, *spectral* and *spatial* features potentially useful for emotion recognition, comparing them to previous proposals through a rigorous experimental evaluation. In particular we assess the effectiveness of various spectral features that were not previously envisaged for the problem of classifying emotions into two-classes: negative versus positive *valence* emotions. Moreover, we study the behaviour of the features considered across different subjects and EEG-electrode locations and examine the possibility of exploiting features from a single-channel EEG setup. The latter point is important as it corresponds to a clear trend in EEG hardware setups for general public applications, which are expected to be maintained as light as possible. Additionally, our evaluation is based on a strict protocol that tests for the generalization capabilities of the system unlike previous works where a clear separation between recordings used in the training set and the ones used in the test set was not always observed [4]. Finally we test various combinations of different feature groups with a view to optimize the recognition accuracy.

The rest of the paper is organized as follows. Section 2 presents a brief discussion of previous works on EEG-based automatic emotion recognition, then the features that are explored in our work are introduced in Section 3. Section 4 describes the dataset used for our evaluation before we expose our results and suggest some conclusions in Sections 5 and 6.

2. RELATION TO PRIOR WORK

EEG-based emotion recognition is a relatively recent research topic that holds a number of difficulties and challenges. One of the major challenges, which is inherent to EEG signal analysis in general, is related to the fact that the recorded signals are easily contaminated by heavy artefacts resulting from various sources of noise, both physiological (especially due to ocular and head-muscle activities) and environmental (due to electromagnetic interference). Consequently EEG recording is usually made under very constrained conditions (often pressing the subjects to remain as steady as possible) which adds complexity to the process of data capturing.

In fact, the availability of quality data that can be used to develop the machine learning algorithms needed to address our problem is critical. Important efforts have been made in this direction [5, 6], though different types of stimuli and emotion annotation strategies were used in different works, making difficult the exploitation of more than one dataset at a time. Among the publicly available datasets, we chose ENTERFACE'06 for our work as it has a number of attractive characteristics (see Section 4).

Consistently with the state-of-the-art in emotion recognition, the *Valence-Arousal* (V-A) representation [7, 8] has been adopted by researchers focusing on EEG-based analysis. Valence represents the

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

state of being reactive to the stimuli, while arousal expresses whether the emotion is rather negative or positive. Following work related to ours [9, 10, 11], we focus on the prediction of the valence value in this study, postponing the analysis along the arousal axis to a future work, in order to reach a better understanding of the specific behaviour of the features considered along each of these dimensions.

A number of recent studies have examined the question of feature selection for EEG-based emotion recognition. Both spatio-temporal and spatio-spectral features have been explored. As far as the former are concerned, the so-called *standard features* [12], corresponding to statistical moments extracted from the temporal waveforms recorded by different electrodes, have proven useful [3, 11]. However spectral features have been more largely used.

The work by Li *et al.* [10] focused on spatio-spectral features where Common Spatial Patterns (CSP) are extracted in the γ band. With a band-selection method and CSP being calculated from the filtered data, the authors performed a binary classification along the valence dimension. Xu *et al.* [4] explored spatio-temporal features, with standard features high order crossings [13, 14], and spectral domain features, among them wavelet features and the energy of spectral components in different frequency sub-bands within the 8-30 Hz range. The authors obtained better results with spectral sub-band features than with standard features. More recently, Jenke [15] focused on electrode and feature selection considering statistics of the signals filtered in the α, θ and β bands, as well as Hjorth parameters [16].

Our study continues the exploration of new features, yet taking an original direction, that is considering features amenable to single-channel EEG-based emotion recognition (essentially alternative spectral features), while comparing them, through a rigorous experimental evaluation, to many of the existing features from previous work (especially as seen in studies considering electrode reduction) [15, 4].

3. FEATURES FOR EMOTION RECOGNITION

Different types of temporal, spectral and spatial features are considered in this study. A first subset of them has been already considered in previous EEG signal analysis tasks, while the second subset is composed of our new proposals, essentially spectral features, which are inspired by research performed in another pattern classification domain where spectral descriptors were extensively studied, namely the audio signal classification domain [17].

3.1. Previously used features

Two types of features can be distinguished here: the ones used in previous work using EEG as a modality for emotion recognition and those exploited in other EEG applications such as Human-Machine Interfaces.

Of the latter, the most commonly used type of features are the so-called *standard* features [12], extracting basic statistical information of the signal, such as the mean, standard deviation, *etc*. These can be considered as limited in the sense that they do not effectively capture useful information about the spectral characteristics of the signals. Hence, researchers have considered using features containing spectral information. They are essentially obtained from the spectrogram of each electrode-signal, computing sub-band power values in either the α, β, θ or γ band.

Alternatively, spatial features have been extensively used, especially the Common Spatial Patterns (CSP) [10, 18, 19]. The technique for extracting CSP designs spatial filters that once applied to the multi-channel EEG signals yield features that maximize the inter-class variance in order to facilitate the discrimination of two classes of EEG signals (in our case the positive end negative valence classes). We refer the interested reader to [20] for more details.

3.2. Proposed features

Among our proposed features, originally described in [17] and never previously used for EEG-signal analysis, we can distinguish different subsets:

- The first, referred to as *Spectral Moments*, is a subset of features based on the first four statistical moments of the EEG-signal magnitude spectra:
  - spectral centroid: $S_c = \mu_1$,
  - spectral width: $S_w = \sqrt{\mu_2 - \mu_1^2}$,
  - spectral asymmetry defined from the spectral skewness: $S_a = \frac{2(\mu_1)^3 - 3\mu_1\mu_2 + \mu_3}{{S_w}^3}$,
  - spectral flatness defined from the spectral kurtosis : $S_k = \frac{-3{\mu_1}^4 - 6\mu_1\mu_2 - 4\mu_1\mu_3 + \mu_4}{{S_w}^4} - 3$,

  where moments μ_i are defined by: $\mu_i = \frac{\sum_{k=0}^{K-1}(f_k)^i a_k}{\sum_{k=0}^{K-1} a_k}$; a_k being the amplitude of the k^{th} component of the Fourier transform of the signal with a frequency of $f_k = \frac{k}{N}$.
- Heuristic spectral shape descriptors:
  - a description of the *spectrum flatness* given by:

    $$SF = \frac{\prod_k a_k^{\frac{1}{K}}}{\frac{1}{K}\sum_k a_k}$$

    and *Spectral Crest Factors* (SCF) defined in each sb sub-band as :

    $$SCF(sb) = \frac{max_{k\epsilon sb} a_k}{\frac{1}{K}\sum_{k\epsilon sb} a_k}$$
  - *Spectral slope* defined as spectral decrease ratio [17]:

    $$S_s = \frac{K\sum_{k=1}^{K} f_k a_k - \sum_{k=1}^{K} f_k \sum_{k=1}^{K} a_k}{K\sum_{k=1}^{K} f_k^2 - (\sum_{k=1}^{K} a_k)^2}$$
  - *Spectral decrease* which is given by [17] :

    $$S_d = \frac{1}{\sum_{k=2}^{K} a_k}\sum_{k=2}^{K}\frac{a_k - a_1}{k-1}$$
  - *Spectral variation* also known as *spectral "flux"* [21]:

    $$S_v = 1 - \frac{\sum_{k=1}^{K} a_k(t-1)a_k(t)}{\sqrt{\sum_{k=1}^{K} a_k(t-1)^2}\sqrt{\sum_{k=1}^{K} a_k(t)^2}}$$
  - *Frequency cutoff* computed as the frequency below which 99% of the total spectrum energy is accounted for.
- Linear Prediction Coding coefficients (LPC) designed as the first coefficients of the filter given by an order 2 Auto-Regressive analysis of the signal.
- Autocorrelation coefficients, corresponding to the first coefficients of the inverse Fourier transform of the signal's periodogram.

Domain	Feature	Description	Frame [s]	Overlap [%]
Temporal	"Standard" features [12]	min, max, skewness, kurtosis, mean, standard deviation, median, mean and max of the absolute values of the first differences, mean and max of the absolute values of the second differences	0.5	50
Frequential	Spectrogram		0.25	50
	Spectral	**Moments, SF, SCF, Cut Off frequency, Autocorrelation, Slope, Variation, Flatness, LPC**		
Spatial	Common Spatial Pattern (CSP) [10]	θ band, α band, γ band, β band, and all the frequencies	0.25	-

Table 1. Features extracted for EEG-based emotion recognition. The features we propose are printed in bold.

4. DATASET AND EVALUATION PROTOCOL

The dataset selected for our experimental study is ENTERFACE'06 [5] which consists of physiological and EEG recordings of 5 male subjects watching images conveying a varying emotional content. Three emotional states of interest are thus considered: *calm*, *positively excited* and *negatively excited*. Each recording corresponds to a 15-minute long session and each session contains 30 blocks, 10 per emotional state. A block is a succession of 5 images corresponding to a single emotional state. 3 such sessions are available for each subject. The whole dataset consists of a total of 90 blocks, hence 450 images.

After each block of 5 images the subject is asked to make a self-assessment of his/her emotional state using the Self-Assessment Manikin (SAM) [7] technique by giving a score between 1 and 5 for valence and arousal components.

The images composing each block are taken from a reference dataset: the International Affective Picture System (IAPS) [22], which is a database of 1196 pictures with a varied emotional content. For ENTERFACE'06 recordings, 3 image subsets were selected, one for each emotional state, according to specific thresholds along the valence and arousal axes. This resulted in a selection of respectively 106, 71 and 50 pictures for the 3 emotional classes considered.

The EEG signals were recorded through 54 channels at a sampling rate of 1024 Hz.

The features are evaluated through the performance of the classifiers exploiting them, where the performance is measured as the valence recognition accuracy. Further, we use a cross-validation procedure designed in such a way to assess the generalisation ability of the recognition system. Xu *et al.* [4] used a 5-fold cross-validation process, *randomly* using 80% of the samples for training and 20% for testing. This approach has the inconvenient not to clearly separate the recordings used to train the classifier and the ones used to provide test data, hence limiting the validity of the results. Therefore we consider a *leave-one-block-out* protocol where data corresponding to one ENTERFACE'06 image-block is used as a test set and the rest of the data is used for training. Two alternative schemes are then evaluated: the *inter-session scheme* where data from the training set contains all the sessions including the one of the block used in the test set; and the *the intra-session scheme* where blocks used in the training set and the ones used in the test set originate from a same session. The results corresponding to each of these situations will be discussed hereafter.

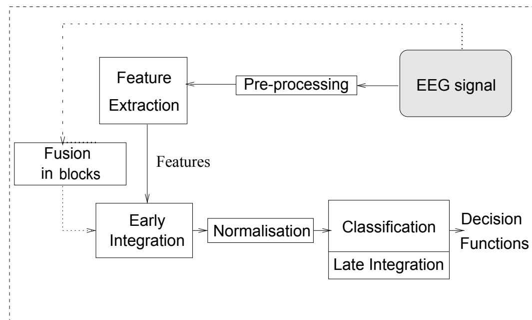

Fig. 1. System Diagram

5. EXPERIMENTS AND RESULTS

In this section, we focus on the different experiments undertaken to study the features presented above.

In all the experiments we use linear Support Vector Machines (SVM). Further, *Early temporal integration* is utilised whereby features extracted from local signal-frames (of sizes given in Table 1) are averaged across the duration of blocks. The classification process, as shown in Figure 1, is thus done for each subject separately using the cross-validation procedure described above.

Features are extracted for each subject across the 54 electrode locations. Table 2 provides, for each subject, the recognition accuracies obtained with the different feature groups, when using features from all electrodes or just from a single one, namely the central Cz electrode. It is important to note that accuracies below 50% correspond to useless recognition systems that do not perform better than chance.

5.1. Subject variability

Consistent with findings of previous works, we observe that the results vary significantly across the different subjects (naturally, individuals react differently to subjective stimuli). Interestingly, two subjects appear to behave in two opposite extreme directions: Subject 5 whose valence values are predicted with high accuracy (up-to 73%) and Subject 1 whose emotional states cannot be properly characterized, yielding scores worse than chance in the majority of feature configurations, and never exceeding 58%.

Comparing the ground-truth labels resulting from the subjects' self-assessment with the ground-truth labels of the corresponding image stimuli (as provided in the IAPS dataset), we find them to be consistent for Subject 5, but not for Subject 1. This tends to confirm that the affective reactions of the latter are quite unusual, hence probably difficult to predict.

	All electrodes						Electrode Cz						CSP		
	Inter-session			Intra-session			Inter-session			Intra-Session			Inter-session		
	Mts	SF	SCF	Mts	SF	SCF	Mts	SF	SCF	Mts	SF	SCF	α	γ	$\alpha + \gamma$
S1	30 %	45 %	48 %	38 %	33 %	32 %	53 %	40 %	28 %	52 %	**58 %**	35 %	38 %	45 %	45 %
S2	48 %	62 %	**65 %**	53 %	53 %	52 %	55 %	52 %	60 %	55 %	42 %	35 %	37 %	55 %	58 %
S3	62 %	53 %	48 %	37 %	48 %	53 %	52 %	63 %	**70 %**	37 %	35 %	33 %	60 %	55 %	60 %
S4	**58 %**	30 %	45 %	35 %	37 %	45 %	48 %	38 %	50 %	40 %	37 %	35 %	40 %	38 %	47 %
S5	70 %	70 %	**73 %**	62 %	60 %	58 %	58 %	67 %	57 %	57 %	50 %	57 %	58 %	53 %	57 %

Table 2. Recognition accuracy (along the valence dimension) for a subset (the best) of the proposed features: Moments (Mts), SF, SCF, and CSP for a dimension of $D_{CSP} = 20$ for each of the 5 subjects (S1 to S5). Best accuracies are printed in bold.

5.2. Comparing features

For the lack of space, only a selection of features resulting in the best overall classification results is retained in Table 2. For instance, CSP features were extracted both on full-band EEG signals and filtered signals in the θ (4-8 Hz), α (8-12 Hz), β (12-30 Hz) and γ (30-+ Hz) bands. Also for each of these versions the reduced CSP dimension was varied in the set $\{2, 4, 20, 40\}$. The best results have thus been obtained with the α and γ bands using 20 CSP coefficients. We notice again a varying behaviour across subjects as different bands (among α and γ) appear to be preferable for different subjects. Further, combining CSP from the two bands does not result in better scores.

Comparing our new spectral feature proposals to previously exploited features, including the different CSP variants, we find the former to outperform the latter in the vast majority of cases. Both the spectral moments, SF and SCF features turn out to be very useful.

In a second experiment, we combined together different pairs of the feature groups presented above. The results are given in Table 3 where only the most effective combinations are presented. When no combination performed better than a single feature group, we kept the latter. It is found that the winning set of features tend to include either spectral attributes or CSP, or both. Examining the detailed results with the aim to better understand the interactions between different features, we notice that Subject 2 responds very well to a specific scheme which consists in combinations of 20 CSP features with Standard features, yielding scores between 76.7% to 78.3%. The best recognition systems for Subjects 3,4 and 5 also essentially rely on combinations with CSP but with respectively LPC Coefficients, spectral moments, and either SF or SCF.

5.3. Generalization ability

In order to assess the generalization abilities of the recognition systems studied, we examine the results obtained using intra-session protocol compared to the inter-session protocol (see Section 4). The results show similar results in intra-session and inter-session protocol, appart for subject 3 whose results are improved considering all sessions for cross-validation. All subjects react differently, and results that are improved by intra-session scheme differe from one subject to another, subject 5 has better results in inter-session scheme, as well as for one feature to another, subject 2 reaching 65 % for SCF and 48 %for spectral moments in considering all the sessions obtains respectively 52 % and 53 % for one session.

5.4. Single-channel setup

In a final experiment, we explore the different features separately in a single-channel EEG setup. In order to do so we select a central electrode, Cz, being less affected by artifacts. The corresponding results for the same spectral features are shown in Table 2. Interestingly, with this setup we have similar results to those obtained with all the electrodes, even reaching 70% accuracy with SCF for subject 3. The single-channel setup is conceivable as most of the necessary information is found by keeping one of the electrodes.

	Score	Features	Parameters
S1	65 %	Spectral power in α band	mean
S2	78 %	CSP β, α + Standards	$D_{CSP} = 20$
S3	70 %	SCF	Central elec. (Cz)
S4	67 %	Spectral Variation	
S5	77 %	CSP γ + SF	$D_{CSP} = 20$ or 40

Table 3. Best score with its associated features for each subject

6. CONCLUSION AND FUTURE WORK

In this paper, we conducted a large number of experiments to study the usefulness of a wide range of EEG-features, including original ones, for a binary classification of emotions along the valence dimension. Special care was taken in using an adequate cross-validation protocol designed to assess the generalisation abilities of the tested recognition systems.

Our results show that the new spectral shape features that we propose are very competitive compared to previously used ones. They are additionally amenable to successful emotion recognition in single-channel setups, which holds a great potential for general public applications.

Future work will consider automatic feature selection algorithms to enable us to efficiently explore the space of feature combinations. Additionally, we will study the behaviour of the features when considering the arousal dimension.

7. REFERENCES

[1] Dacher Keltner, Paul Ekman, Gian C Gonzaga, and Jennifer Beer, "Facial expression of emotion," *Oxford University Press*, 2003.

[2] Björn Schuller, Anton Batliner, Stefan Steidl, and Dino Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Speech Communication*, Feb. 2011.

[3] Guillaume Chanel, Julien Kronegg, Didier Grandjean, and Thierry Pun, "Emotion assessment: Arousal evaluation using

EEGs and peripheral physiological signals," *Multimedia Content Representation, Classification and Security*, pp. 530–537, 2006.

[4] Haiyan Xu and Konstantinos N. (Kostas) Plataniotis, "Affect recognition using EEG signal," *IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 299–304, Sept. 2012.

[5] Arman Savran, Koray Ciftci, Guillame Chanel, and J Cruz Mota, "Emotion detection in the loop from brain signals and facial images," 2006.

[6] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[7] Margaret M. Bradley and Peter J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.

[8] Albert Mehrabian and James A Russell, *An approach to environmental psychology*, The MIT Press, 1974.

[9] Robert Horlings, Dragos Datcu, and Leon J.M. Rothkrantz, "Emotion recognition using brain activity," in *International Conference on Computer Systems and Technologies - CompSysTech'08*, 2008.

[10] Mu Li and Bao-Liang Lu, "Emotion classification based on gamma-band EEG," *Conference Proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1323–1326, 2009.

[11] Zahra Khalili, "Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG," *Proceedings of International Joint Conference on Neural Networks (IJCNN 2009).*, pp. 1571–1575, 2009.

[12] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[13] Benjamin Kedem and Michael Frey, "Time Series Analysis by Higher Order Crossings," *IEEE press Piscataway*, May 1994.

[14] Panagiotis C. Petrantonakis and Leontios J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, 2010.

[15] Robert Jenke, Angelika Peer, and Martin Buss, "Effect-Size-Based Electrode and Feature Selection for Emotion Recognition from EEG," 2013.

[16] Bo Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[17] Geoffroy Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," Tech. Rep., IRCAM, 2004.

[18] Sander Koelstra, Ashkan Yazdani, Mohammad Soleymani, and C Mühl, "Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos," *Brain informatics*, vol. 6334, pp. 89–100, 2010.

[19] Irene Winkler, M Jäger, V Mihajlović, and T Tsoneva, "Frontal EEG asymmetry based classification of emotional valence using common spatial patterns," *World Academy of Science, Engineering and Technology*, vol. 45, pp. 373–378, 2010.

[20] Benjamin Blankertz and Ryota Tomioka, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. XX, pp. 1–12, 2008.

[21] Eric Scheirer, Malcolm Slaney, and Palo Alto, "Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1331–1334, 1997.

[22] Peter J. Lang, Margaret M. Bradley, and Bruce N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," *Technical Report A-8*, 2008.