Ranking of EEG Time-domain Features on the Negative Emotions Recognition Task

Firgan Nihatov Feradov and Todor Dimitrov Ganchev

Abstract – **The accuracy of automated emotion recognition depends on the quality of EEG signal descriptors. In the present contribution we report on an experimental evaluation of ten time-domain EEG signal descriptors with respect to their applicability to the task of negative emotions recognition. The ranking of these descriptors based on their estimated practical worth shows that the mean of the absolute values of the first difference of the normalized signal contributes for the highest recognition accuracy.**

Keywords – **Emotion recognition; Electroencephalography; EEG signals; Feature ranking.**

I. INTRODUCTION

In the last decade there is an increasing interest towards the detection of emotions from EEG signals. This is mainly due to the high demand for intelligent human-machine and brain-computer interaction interfaces, which are important part of information support, health care, and educational training applications.

Nowadays human-machine interaction technology is dominated by the statistical machine learning paradigm, which assumes the existence of datasets representative to the operational conditions of a certain application. These datasets are used for the creation of statistical model(s) representing each category of interest, or for the creation of discriminative classifiers. Given sufficient amount of representative data, the classification accuracy depends on the discriminative power of the classifier, but more importantly on the informative value of the signal descriptors fed to the classifier.

In the present work we carry out an experimental evaluation of various signal descriptors, based on the time domain EEG signal, and evaluate their applicability to the task of automated recognition of negative emotional states. In particular each descriptor from two widely-used feature sets is compared with its counterparts and with the short-time energy of an EEG signal, used as a feature [1-3].

Specifically, the first set of EEG features is the set designed by Bo Hjorth [4]. It consists of three features (*Activity*, *Mobility*, and *Complexity*) describing the amplitude, frequency, and shape of an EEG signal. In brief, *Activity* is defined as the amplitude variance of a signal. It is considered to have the necessary additive property to allow integration of different observations during the epoch into one representative figure. When computed the *Activity* has a high or low value if the high

F. Feradov is with the Faculty of Electronics Engineering, Technical University of Varna, str. Studentska 1, 9010 Varna, Bulgaria, e-mail: firgan.feradov@tu-varna.bg

T. Ganchev is with the Faculty of Electronics Engineering, Technical University of Varna, str. Studentska 1, 9010 Varna, Bulgaria, e-mail: tganchev@tu-varna.bg

frequency components in the signal are few or respectively many. *Activity* corresponds to spectral analysis in the frequency domain. *Mobility* is calculated as the square root of the ratio between the variances of the first derivative and *Activity*, and represents the average of the frequency of the signal. This descriptor corresponds to the calculation of the standard deviation of a signal in the frequency domain. *Complexity* is a measure of details with reference to the curve shape of the sine wave.

Activity, *Mobility*, and *Complexity* were initially designed for the creation of EEG-based human-computer interfaces, but later on they have been proven universal and applicable to other tasks that rely on EEG signals, including the detection of emotional states from EEG signals [5].

The second set of features evaluated here was proposed by Picard et al. [6], who compared multiple algorithms for feature-based recognition of emotional states on a given set of data. Among these are six statistical descriptors, such as: (i) the means of the raw signal, (ii) the standard deviation of the raw signal, (iii) the means of the absolute values of the first and second differences of the raw signal, and (iv) the means of the absolute values of the first and second differences of the normalized signal. These descriptors were purposely designed for the emotion detection task and account for the physiological activity of the body and brain. These six descriptors were employed in related studies on emotion detection from EEG signals [7-9].

In Section II we outline the ten EEG descriptors evaluated in the present work. In Section III we describe the common experimental setup used in the feature performance evaluation study. In Section IV we provide details on the estimation of person-specific thresholds for the recognizer of negative emotional states. The experimental results are reported in Section V, and in Section VI we provide summary and concluding remarks.

II. FEATURE EXTRACTION

The EEG signal descriptors considered here are characterized with low complexity and all features are derived directly from the time-domain signal. In brief, first the EEG signal $x_i(n)$ is pre-processed for reducing the interference from muscular and eye movement activity, which results in the free of artefacts EEG signal $\hat{x}_i(n)$. Here the subscript i stands for the channel number. All channels are processed uniformly so in further discussion we drop the index i but it reminds implied. All features are computed for short frames of the EEG signal, obtained through a sliding window of 343.75 milliseconds which moves with a skip rate of 85.9 milliseconds. Successive frames overlap with 75%. Therefore, the total number of

successive overlapping frames in a recording with N samples is:

$$P = fix\left(\frac{N-K+L}{L}\right), \qquad (1)$$

where the operator *fix* stands for rounding towards the smaller integer number, L is the predefined step size in samples, and K is the frame size, also in samples. Next, for each frame we compute the EEG signal feature that is going to be evaluated, as follows:

a) Short-time energy:

$$E_p = \sum_{k=1}^{K} \widehat{x}(k)^2, \ p = 1,2,...,P. \qquad (2)$$

b) Mean value:

$$\mu_p = \frac{1}{K}\sum_{k=1}^{K} x(k), \ p = 1,2,...,P. \qquad (3)$$

c) Standard deviation:

$$\sigma_p = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(\widehat{x}(k)-\mu_p)^2}, \ p = 1,2,...,P. \qquad (4)$$

d) The mean of the absolute values of the first difference of the signal (MAVFDS):

$$\delta_p = \frac{1}{K-1}\sum_{k=1}^{K-1}|\widehat{x}(k+1)-\widehat{x}(k)|, \ p = 1,2,...,P. \qquad (5)$$

e) The mean of the absolute values of the second difference of the signal (MAVSDS):

$$\gamma_p = \frac{1}{K-2}\sum_{k=1}^{K-2}|\widehat{x}(k+2)-\widehat{x}(k)|, \ p = 1,2,...,P. \qquad (6)$$

f) The mean of the absolute values of the first difference of the normalized signal (MAVFDNS):

$$\widetilde{\delta}_p = \frac{1}{K-1}\sum_{k=1}^{K-1}|\widetilde{x}(k+1)-\widetilde{x}(k)| = \frac{\delta_p}{\sigma_p}, \ p = 1,2,...,P. \qquad (7)$$

g) The mean of the absolute values of the second difference of the normalized signal (MAVSDNS):

$$\widetilde{\gamma}_p = \frac{1}{K-2}\sum_{k=1}^{K-2}|\widetilde{x}(k+2)-\widetilde{x}(k)| = \frac{\gamma_p}{\sigma_p}, \ p = 1,2,...,P. \qquad (8)$$

h) Activity:

$$Activity = \sigma_p^2 \qquad (9)$$

i) Mobility:

$$Mobility = \frac{\sigma_d}{\sigma_p} \qquad (10)$$

j) Complexity:

$$Complexity = \frac{\sigma_{dd}/\sigma_d}{\sigma_d/\sigma_p} \qquad (11)$$

Next, statistical standardization of the calculated descriptors was performed, so that their distributions are normalized to zero mean value and unit standard deviation:

$$Z = \frac{D_p - \mu_p}{\sigma_p} \qquad (12)$$

The mean value μ_p and the standard deviation σ_p are estimated for the feature of interest D_p and the process is repeated for each of the EEG channels. The subscript p is the frame index.

III. EXPERIMENTAL SETUP

The evaluation of the aforementioned ten EEG signal descriptors was carried out following a common experimental protocol, based on the Database for Emotion Analysis using Physiological signals (DEAP) [15]. All descriptors were computed from the same set of EEG signals. Each descriptor was used separately to create a classifier, which were tested on another dataset. All descriptors were ranked according to the recognition accuracy.

The DEAP database consists of recordings from thirty-two participants with a total of 40 recordings per participant. Each recording, made while the participants were watching musical video-clips, consists of 40 channels. These include 32 EEG channels, electromyographic (EMG), electrooculographic (EOG) and other channels, all taken from different parts of the head and the body. The dataset includes original and pre-processed recordings from these forty channels. Frontal face videos and detailed metadata for the participants are also included in the database. All EEG recordings were self-annotated by the subjects participating in the data collection.

The EEG recordings of 10 participants (with numbers № 2, 11, 17, 21, 22, 24, 28, 29, 30, 32) were selected for our experiments, based on the annotations of the EEG songs. We aimed at balance between the numbers of songs tagged as negative and positive for each participant. Each participant's data was split in three parts – training, development, and testing dataset, which consist of 20%, 20%, and 60% of the available recordings. However, depending on the distribution of song ratings for each participant, these percentages varied up to 5%. This led to roughly 8 to 9 recordings used for training, 8 to 9 recordings used for development, and 22 to 24 recordings used for testing, for each participant. The total number of test recordings is N_{rec}=227.

The split of recordings into these three datasets was performed in the following way: The dataset of each participant was split into two groups – *negative* and *non-negative* – depending on the like/dislike rating of the recording. Each rating provided in the database indicates the personal preferences of the participant. The like/dislike rating's range is from 1 to 9, where 1 corresponds to the lowest rating (*disliked*) and 9 is the highest rating – *liked*. In the current experiment, recordings which had like/dislike rating of 4 or lower were tagged as *negative*, while recordings with rating higher than 4 were tagged as *non-negative*. In few cases when the number of definitely tagged recordings was not sufficient a 5% tolerance was applied to the separation threshold. All EEG recordings for the selected participants were used during the experiments, which totals to 400 EEG recordings.

The classification performance of the trained models is evaluated in terms of percentage correct detections:

$$correct = \frac{H}{N_{rec}} \times 100, [\%] \qquad (13)$$

In all experiments we trained an SVM classifier with radial basis function kernel. In order to find out the optimal training parameters for every emotion classifier, a series of grid searches were implemented on the development dataset. Grid searches were carried out for each of the individual descriptors, for each participant. In particular, the adjustable parameters of the SVM tuned during the grid search were:

- *boxconstraint* – the box constraint C for the soft margin, where C is a positive numeric scalar or a vector.
- *rbf_sigma* – a positive number specifying the scaling factor in the Gaussian radial basis function kernel.

The grid searches ranged between 0.1 and 100 for both parameters. The optimal values found during the grid search were used during the evaluation of each detector. A total of 200 grid searches were conducted.

IV. DETECTION OF NEGATIVE EMOTIONS

The evaluation of the features is conducted through the use of a detector of negative emotional states, presented in [1], which is based on the following principle:

The training data set, composed of the most indicative examples of negative and non-negative recordings, was used to train a person-specific SVM classifier. The model, generated this way, was then tuned on the development dataset, so that a person-specific decision threshold Tr can be computed:

$$Tr = \frac{\frac{1}{n}\sum_{f=1}^{n} D_{neg,f} + \frac{1}{m}\sum_{l=1}^{m} D_{pos,l}}{2}, \qquad (14)$$

where D_{neg} and D_{pos} are the portions of development data consisting of n recordings with negative tags and m with non-negative tags (neutral or positive).

During the evaluation of each classifier, the threshold (14) was used for making a decision for each recoding. The person-specific recognition accuracy for each individual descriptor was computed based on the results for all recordings.

V. EXPERIMENTAL RESULTS

Based on the common experimental setup, described in Section III, we performed a comparative study of the ten EEG descriptors. The experimental results are presented in Tables 1 and 2. In Table 1 we present the averaged classification accuracy for each individual descriptor and in Table 2 the results per participant.

In Table 1 a comparison between the recognition results, observed for each feature is shown. The first column shows the feature name, whereas the features are ordered in descending order with respect to their performance – the feature that had highest average accuracy is placed first, while the feature with lowest average accuracy is respectively placed last, on the bottom of the table. The second column shows the highest

classification accuracy, observed for each feature, while in the third column the lowest achieved accuracy is shown. In the fourth column we present the average accuracy observed for each feature. In the fifth and final column we can see the percentage of successful classifications. This comparison is made, because in some cases the trained models were not able to correctly classify the signals and reach a solution for the given task.

TABLE 1. CLASSIFICATION ACCURACY ACHIEVED BY THE FEATURES

Feature name	Max accuracy	Min accuracy	Mean accuracy	Successful classification
MAVFDNS	87.0 %	59.1%	68.7 %	90 %
MAVSDNS	82.6 %	58.3%	68.6 %	80 %
Mobility	75.0 %	59.1 %	67.0 %	90 %
MAVSDS	73.9 %	59.1 %	66.6 %	100 %
Complexity	78.3 %	59.1 %	66.0 %	90 %
Energy	81.8 %	58.3 %	65.7 %	100 %
Activity	70.8 %	59.1 %	64.0 %	80%
MAVFDS	68.2 %	54.5 %	63.8 %	100 %
St. deviation	68.2 %	59.1 %	63.4 %	100 %
Mean value	72.7 %	59.1 %	62.9 %	90%

As presented in Table 1, we observed the highest classification accuracy for the mean of the absolute values of the first and second differences of the normalized signal (MAVFDNS and MAVSDNS), equal to 68.7% and 68.6%. Overall, the more complex of the six statistical features displayed higher classification capabilities, while the simpler statistical features were not as descriptive, with *Mean value* and *Standard deviation* having the lowest classification accuracy (63.4% and 62.9%) among all descriptors considered here. A performance division, based on the complexity of the feature can also be observed in Hjorth's set of features, where *Complexity* and *Mobility*, also performed well with mean accuracy of 67% and 66% respectively, while the mean accuracy of the *Activity* parameter was 64%.

In Table 2 we present detailed information about the recognition accuracy observed for each individual descriptor, on the dataset of the particular participant. The cases, in which the classifier failed detect the data correctly are marked with "X".

Although most of the examined features showed similar performance, some variations did exist. One such example are Hjorth's descriptors, which were computationally expensive but led to an increase in the recognition accuracy. Another case where difference in performance was observed was during classification with *Mean value* as a descriptor. This feature showed performance inconsistencies and the recognition accuracy varied greatly, depending on the grid searched parameters, used for the creation of the SVM classifier.

VI. CONCLUSIONS

A study of the applicability of ten time-domain features and their informative value with respect to classification accuracy capabilities was carried out on the DEAP dataset. We ranked these features with regard to the recognition accuracy.

TABLE 2. CLASSIFICATION ACCURACY OF THE EVALUATED DESCRIPTORS PER PARTICIPANT.

Features:	*Par. 02*	*Par. 11*	*Par. 17*	*Par. 21*	*Par. 22*	*Par. 24*	*Par. 28*	*Par. 29*	*Par. 30*	*Par. 32*
Mean Value	66.3 %	68.2 %	59.1 %	60.9 %	X	72.7 %	62.5 %	56.5 %	60.9 %	59.1 %
St. Deviation	61.6 %	59.1 %	68.2 %	60.9 %	68.2 %	59.1 %	62.5 %	65.2 %	65.2 %	63.6 %
MAVFDS	61.6 %	54.5 %	68.2 %	65.2 %	68.2 %	63.6 %	66.7 %	56.5 %	69.6 %	63.6 %
MAVFDNS	66.3 %	X	68.2 %	65.2 %	68.2 %	63.6 %	62.5 %	78.3 %	87.0 %	59.1 %
MAVSDS	66.8 %	63.6 %	72.7 %	65.2 %	72.7 %	68.2 %	62.5 %	60.9 %	73.9 %	59.1 %
MAVSDNS	61.6 %	63.6 %	72.7 %	X	63.6 %	68.2 %	58.3 %	82.6 %	78.3 %	X
Activity	56.3 %	59.1 %	59.1 %	X	68.2 %	X	70.8 %	65.2 %	65.2 %	68.2 %
Mobility	61.6 %	63.6 %	68.1 %	X	63.6 %	68.2 %	75.0 %	73.9 %	69.6 %	59.1 %
Complexity	61.6 %	68.2 %	59.1 %	X	63.6 %	68.2 %	62.5 %	78.3 %	73.9 %	59.1 %
Energy	58.3 %	59.1 %	63.6 %	69.6 %	68.2 %	81.8 %	66.7 %	60.9 %	69.6 %	59.1 %

The highest average accuracy was achieved with the mean of the absolute values of the first differences of the normalized signal (MAVFDNS) – 68.7%. The feature with the lowest average accuracy was mean value – 62.9%. Also, the observed results indicate an absolute increase in recognition accuracy with 0.9%, when compared to the results reported in previous related work [2].

The future research will aim at improvement of the recognition accuracy of negative emotions from EEG signals, based on combinations of the most discriminative descriptors.

ACKNOWLEDGEMENT

REFERENCES

[1] Feradov, F., Ganchev, T. "*Detection of Negative Emotional States from Electroencephalographic (EEG) signals*", Annual Journal of Electronics, vol. 8, pp. 66-69, 2014.

[2] Feradov, F., Ganchev, T, "*Detection of Negative Emotions from EEG Signals using Time-Domain features*", Proc. of the International Scientific Conference UNITECH-2014, Gabrovo, vol. 1, pp. 315-320, 2014.

[3] Feradov, F., Ganchev, T. "*Evaluation of Electrode Setups for EEG-Based Automated Emotion Detector*", Proc. of Technical University – Varna, 2014.

[4] Hjorth, B., "*EEG analysis based on time domain properties, Electroencephalography and Clinical Neurophysiology*", ISSN 0013-4694, vol. 29, no. 3, pp. 306-310, September 1970.

[5] Seung-Hyeon Oh, Yu-Ri Lee, and Hyoung-Nam Kim "*A Novel EEG Feature Extraction Method Using Hjorth Parameter*", International Journal of Electronics and Electrical Engineering, vol. 2, no. 2, pp. 106-110, June 2014

[6] Picard, R.W., Vyzas, E., Healey, J. "*Toward Machine Emotional Intelligence: Analysis of Affective Physiological State*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp.1175-1191, Oct. 2001.

[7] Yuen, C.T., San, W.S., Seong, T.C., Rizon, M. "*Classification of Human Emotions from EEG Signals using Statistical Features and Neural Network*". International Journal of Integrated Engineering, vol. 1, no. 3, 2009.

[8] Petrantonakis, P.C., Hadjileontiadis, L.J. "*Emotion Recognition From EEG Using Higher Order Crossings,*" IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp.186-197, March 2010.

[9] Wang, X-W., Nie, D., Lu, B-L. "*EEG-Based Emotion Recognition Using Frequency Domain Features and Support Vector Machines*", Proc. of the 18th International Conference Neural Information Processing, ICONIP-2011, Shanghai, China, November 13-17, 2011, part I, pp. 734-743.

[10] Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I. *DEAP: A Database for Emotion Analysis using Physiological Signals*, IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18-31, 2012.