Emotion Recognition using cvxEDA-Based Features

H. Ferdinando[1,2], E. Alasaarela[1]
[1]*Health and Wellness Measurement Group, Optoelectronics and Measurement Technique Unit, University of Oulu, Finland.*
[2]*Department of Electrical Engineering, Petra Christian University, Indonesia.*
hany.ferdinando@oulu.fi

***Abstract*—The MAHNOB-HCI database provides baselines for several modalities but not all. Up to now, there are no such baselines using EDA signal for valence and arousal recognitions. Because EDA is one of the important signals in affect recognition, it is necessary to have baseline accuracy using this signal. Applying cvxEDA, EDA tool analysis based on convex optimization, to GSR signals resulted phasic, tonic, and sudomotor neuron activity (SMNA) phasic driver. There were two sets of features extracted, i.e. features from stimulated stage only and ratio of features from stimulated to relaxation stages in addition to the former set. Using kNN to solve the 3-class problem, the best accuracies under subject-dependent scenario were 74.6 ± 3.8 and 77.3 ± 3.6 for valence and arousal respectively while subject-independent scenario resulted in 75.5 ± 7.7 and 77.8 ± 8.0 for valence and arousal correspondingly. Validation using LOO gave 75.2% and 77.7% for valence and arousal respectively. cvxEDA method looked promising to extract features from EDA as the results were even better than the best results in the original database baseline. Some future works are using other feature extraction method, enhancing the accuracies by employing supervised dimensionality reduction and using other classifiers.**

***Index Terms*—cvxEDA; EDA; Emotion Recognition; MAHNOB.**

I. INTRODUCTION

Previous studies [1]–[4] on emotion recognition using the MAHNOB-HCI database for affect recognition provided an important contribution to research on affective computing. Soleymani et al. [1] provided some baseline accuracies for several modalities using the signal from the MAHNOB-HCI database for the 3-class problem in valence (it represents the degree of pleasantness) and arousal (it represent the state of being awoken). Providing baselines from all possible combinations, including individual signal, however, seemed impossible, such that only some important combinations were provided. The best performances were 68% and 76% for valence and arousal respectively by combining features from EEG and eye gaze. Ferdinando et al. [2] provided baseline accuracy for valence, 43%, and arousal, 48%, using ECG signals only and they were gradually improved by employing different methods. To the best of the author's knowledge, the best baseline accuracies using ECG signals only were 64% and 66% for valence and arousal respectively [4], validated under the subject-dependent scenario. The ones under subject-independent scenario were 59% [3] and 70% [4] for valence and arousal correspondingly.

Another interesting modality for affect recognition is electrodermal activity (EDA) signal. Kreibig wrote that Autonomic Nervous System (ANS) activities related to emotion regulation emerged in cardiovascular and skin conductance (SC) measurement [5], bringing the fact that SC cannot be neglected in emotion recognition research. This signal has been the subject of many classic studies in emotion recognition in the absence of the other signals [6]–[9]. To the best of the author's knowledge, there is no baseline accuracy using EDA measurement from the MAHNOB-HCI while EDA has shown good results in other research and this paper aims to fill this gap.

There are many methods to analyze EDA signals, e.g. Ledalab with Continuous Decomposition Analysis (CDA) [10] and Discrete Deconvolution Analysis (DDA) [11], EDA Explorer [12], cvxEDA [13]. EDA Explorer is available in Python as well as in web-based application while the others are available in Matlab, by which the experiments in this paper were carried out.

Greco et al. [9] compared Ledalab-based model and cvxEDA-based model for emotion recognition problem and found that the latter outperformed the former. For this reason, this paper used cvxEDA to analyze EDA signals from the MAHNOB-HCI database.

This study also aims to contribute to research on effective computing utilizing physiological signals, especially EDA, by applying cvxEDA to extract useful features for the classifier to solve the 3-class problem in valence and arousal. The results complement the existing baseline accuracies for the MAHNOB-HCI database [1], [3], [4].

II. LITERATURE STUDIES

A. Emotion Recognition using EDA

EDA is defined as automatic changes in the skin electrical properties due to sweat gland activities [9]. The sweat glands secretion depends on many factors and one of them is emotion stimulation. The sweat on skin influence the conductivity of the skin and researchers uses this quantified property to recognize emotion.

Gunes and Hung [14] wrote that using the facial expression for emotion recognition would come to a dead end unless multimodality was used and one of the proposed biosignals was EDA. Facial expression emotion recognition requires cameras, which were not practical in most situation, while EDA is measured non-invasively on fingers, wrist, or hand palm, using simple circuits. Moreover, the EDA, which is controlled by ANS, corresponding to arousal state of human being [15].

Lanata et al. [6] used EDA glove based on textile-integrated electrodes to discriminate affective states of 35 subjects stimulated by IAPS images. It achieved promising results to separate neutral and 4-level of arousal using standard, including frequency-domain features, and nonlinear features. Of note, nonlinear features usually are computationally high cost.

Yang and Liu [8] extracted nonlinear statistic features after surrogate data analysis from EDA to separate five discrete emotions using several classifiers with accuracy up to 87%. Results from this study provide strong evidence that relationship between EDA signal and emotion is nonlinear.

Greco et al. [9] used cvxEDA to analyze EDA and then extract useful features for emotion recognition and mood/mental disorder assessment. Specific for emotion recognition experiment where subjects were stimulated by IAPS images, features from cvxEDA analysis was superior to the ones from Ledalab analysis to separate 4-class of arousal, 72% to 37%, while both had the same performance on valence. The cvxEDA looked promising for EDA signal analysis.

B. *The cvxEDA*

The cvxEDA is a method to analyze EDA using convex optimization, proposed by Greco et al. [13], which casts the EDA deconvolution as a quadratic optimization problem. The EDA generation was modeled based on the following assumptions:

- Skin conductance response (SCR) is preceded by burst, generated by a sparse and nonnegative neural signal, from sudomotor nerves controlling the sweat glands.
- The number of recruited sweat glands and the amplitude of a firing burst have a linear relationship, which makes the time course of a single SCR induced by a neural burst is free from the previous ones although their SCRs overlap, in other words, it is LTI.
- The sweat diffusion process has a relatively stable subject-specific impulse response filter (IRF) for all SCRs from the same subject.
- The phasic activity is superimposed to a slowly varying tonic activity with spectrum below 0.05 Hz.

The cvxEDA splits data into phasic (r), tonic (t), and a noise component (ε), such that the observation model (y) can be written as

$$y = r + t + \varepsilon \tag{1}$$

within r, the shape of a single phasic response is modeled using Bateman function,

$$h(\tau) = (e^{-\tau/\tau_0} - e^{-\tau/\tau_1})u(\tau) \tag{2}$$

where τ_0 and τ_1 are the slow and fast time constant, and $u(\tau)$ is a unit step function. By representing ARMA model of Equation (1) into cascade ARMA,

$$r = MA^{-1}p \tag{3}$$

where p represents sudomotor neuron activity (SMNA) and M and A are a tridiagonal matrix, see detail in [13].

Tonic component, t, is represented as the cubic B-spline basis function

$$t = B\lambda + Cd \tag{4}$$

where B is a matrix for cubic B-spline basis function, λ is the vector of spline coefficient, C is an $Nx2$ matrix with $C_{i,1} = 1$ and $C_{i,2} = 1/N$, d is a $2x1$ vector with offset and slope coefficients for the linear trend.

The observation model can be rewritten by substituting Equation (3) and (4) into (1). The goal is to identify maximum a posteriori (MAP) spike train of p and t, parameterized by $[q, \lambda, d]$, for the measured EDA signals.

The SMNA p is modeled using Poisson distribution, but later an exponential distribution of the same mean replaces it to keep the analysis tractable. For tonic component, $\Delta = 10s$ is used to make the sampling frequency is exactly twice the upper band limit, i.e. 0.05 Hz and assumes that vector λ has normal distribution so does the noise term, ε. These form likelihood term for each q, λ, and d. By substituting these terms into

$$P(q,\lambda,d \mid y) \propto P(y \mid q,\lambda,d)P(q)P(\lambda)P(d) \tag{5}$$

and taking the logarithm of Equation (5), the optimization problem representing the cvxEDA algorithm is

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|Mq + B\lambda + Cd - y\|_2^2 + \alpha\|Aq\|_1 + \frac{\gamma}{2}\|\lambda\|_2^2 \\ &\text{subject to } Aq \geq 0 \end{aligned} \tag{6}$$

III. Materials and Methods

A. *The Database*

The MAHNOB-HCI database for affect recognition involved 30 participants stimulated emotionally by pictures and fragments of movies [1]. The database includes the following synchronized signals: 32-channel EEG, peripheral physiological signals (ECG, temperature, respiration, and skin conductance), eye gaze, face and body camera, and audio. For emotion elicitation experiment, the protocol marked 30 seconds before and after the stimulated session for relaxation with a single pulse to separate them.

The experiments in this paper used data downloaded from the database server under “Selection of Emotion Elicitation”. For each session, baselines (measured during relaxation stage) and response (measured stimulated stage) signals were separated to each other based on single marking pulses. The relaxation signals used in this experiment were the ones before the stimulated stage only as the changing from relaxation to stimulated may provide good separation among the classes.

B. *Feature Extraction*

The EDA signals, sampled at 1024 Hz but downsampled to 256 Hz to save storage space [1], were taken from channel GSR3 of the database. Although EDA and emotion have a nonlinear relationship [8], extracting nonlinear features is usually time-consuming, such that it was not used in this study. On the other hand, the cvxEDA can be applied directly to raw signal, making this method even more interesting.

Before applying the cvxEDA, which requires signals in Siemens, the original signals in Ohm were converted to Siemens. Following this conversion, three signals were extracted using cvxEDA, i.e. phasic (r), tonic (t), and sparse SMNA of phasic component (p) using default parameters. On completion of cvxEDA analysis, features were extracted from the phasic (r), tonic (t), and sparse SMNA of phasic component (p).

The extracted features were the time-domain features as proposed by Greco et al. [9], statistical distribution as applied by Ferdinando et al. [3] in ECG signal analysis, and frequency domain analysis. Mostly, the number of significant SCR are calculated within the 5-second window after the stimulation, but the EDA signals from the MAHNOB-HCI do not have the same length. This brought consequences to modify the definition of this index.

Several features were extracted:

- nSCR1, number of significant SCR within 5-second non-overlap window, divided by number of the window.
- nSCR2 = number of significant SCR within 5-second non-overlap window, divided by length of the signal in seconds.
- nSCR3 = number of significant SCR divided by the length of the signal in seconds.
- The area under the curve (AUC) of phasic and tonic signals.
- 14 items of statistical distribution: mean, standard deviation, Q1, median, Q3, IQR, percentile 2.5, percentile 10, percentile 90, percentile 97.5, maximum, skewness, and kurtosis.

Two sets of features were used in this study, i.e. features from the stimulated stage only (31 features) as *feaure1* and ratio of features in stimulated to relaxation in addition to the first set (62 features) *feature2*. Involving ratio of stimulated to relaxation stages as features should give larger discriminant values to separate certain class from the other because when someone was stimulated the SMNA generated sparse and non-negative neural signal to initiate burst that later emerged as SCR, see the first assumption in the discussion about the cvxEDA. as they were not present during the relaxation stages. Frequency-domain features, Power in 0–0.1 Hz, 0.1–0.2 Hz, 0.2–0.3 Hz, 0.3–0.4 Hz, [6] were also added to both *feature1* and *feature2* to get other feature set *feature3* (42) and *feature4 (84)*. Finally, a sequential forward floating search algorithm was applied to select a set of features offered large discriminant value. This procedure, however, also acted as a dimensionality reduction process.

C. Classifier and Validations

A kNN classifier was used to solve the 3-class problem in valence and arousal [4]. The results were validated under subject-dependent and subject-independent scenarios as in [3], [4]. For subject-dependent scenario, 20% of the samples were held out for validation, while the rest were subject to 10-fold cross validation, repeated for 1000 times with new resampling for each repetition. Reported accuracies were the average over the repetition. Standard deviations were also provided to evaluate variation among the repetition.

Subject-independent scenario evaluates if the proposed features are ready for a general system where the classifier recognizes emotion based on new samples. At first, all samples belong to certain participant were excluded for testing while the rest were used to build the model. This process was repeated for all subjects. The reported accuracy was the average. This validation was called leave-one-subject-out (LOSO) validation [3].

Validation using leave-one-out (LOO) method was also used to compare the results with the ones from Soleymani et al. [1] which used LOO also. LOO is excluding one sample as test while using the rest of the sample to make a model.

The results were compared to the ones from Soleymani et al. [1] and Ferdinando et al. [3], [4] using t-test with significance level 0.05. The same test was also used to select the best performance among all.

IV. Results and Discussions

Frequency-domain features in *feature3* had no contribution since the sequential forward floating search process for *feature1* and *feature3* produced the exact same results, while the results from *feature2* and *feature4* were different. These facts brought consequences that *feature3* was discarded from this study.

Table 1 summarizes experiments for both valence and arousal in subject-dependent and subject-independent scenarios. Results in Table 1 were compared to ECG-based features in Ferdinando et al. [3], [4]. Apparently, they all outperformed ECG-based features performance, shown by very small p-values from t-test, except for *feature2* in arousal in the subject-independent scenario. These findings were remarkable as features from ECG requires exhausted computation, i.e. feature selection and all experiment with kNN were done for features from each combination of window size and overlap parameters in spectrogram analysis [3]. On the other hand, utilizing cvxEDA only requires four combinations, including *feature3*, see section III-C.

Table 1
Accuracies of all experiments written in mean and standard deviation for valence and arousal, validated using subject-dependent and subject-independent scenarios

Feature sets	Subject-dependent		Subject-independent	
	Valence (%)	Arousal (%)	Valence (%)	Arousal (%)
feature1	67.1 ± 4.3	77.2 ± 3.9	68.9 ± 9.4	77.6 ± 8.9
feature2	69.9 ± 4.1	69.4 ± 4.2	72.4 ± 8.3	70.9 ± 10.4
feature4	**74.6 ± 3.8**	**77.3 ± 3.6**	**75.5 ± 7.7**	**77.8 ± 8.0**

Table 2 presents results from LOO validation for both valence and arousal. It was evident that results from *feature4* outperformed the ones accomplished by Soleymani et al. [1]. Moreover, these were achieved using only single modality.

Table 2
Accuracies of all experiments written for valence and arousal validated using LOO

Feature sets	Valence (%)	Arousal (%)
feature1	68.4	77.7
feature2	71.3	71.1
feature4	**75.2**	**77.7**

Summary of all significance tests is shown in Table 3. It was apparent that mostly the EDA-based features outperformed the other, see checked column.

Table 3
Summary of t-test for significant difference to Ferdinando et al. [3], [4] and Soleymani et al. [1]

Feature sets	Ferdinando et al. [3], [4]				Soleymani et al. [1], LOO	
	Subject-dependent		Subject-independent			
	V	A	V	A	V	A
feature1	✓	✓	✓	✓	✓	✓
feature2	✓	✓	✓		✓	
feature4	✓	✓	✓	✓	✓	✓

Among *feature1*, *feature2*, and *feature4*, it was obvious that *feature4* demonstrated its superiority within all validations method in most cases. The significance test confirmed this finding as the p-values for all comparisons were close to zero, indicating significant differences. Involving ratios between stimulated and relaxation stages as features contributed larger discriminant among the class such that it improved the performance. To be more specific, a ratio based on frequency-domain features contributed more than the ones from standard EDA feature and statistical distribution.

LOSO validation through subject-independent scenario exposed interesting results. Building a model by excluding all samples from one subject for testing, degrade performances were expected but Table 1 exposed that the performances were close to the ones from the subject-dependent experiment, only the variation among the experiments were relatively higher, which could not be avoided. This is a surprising and unexpected result.

Turning now to confusion matrices for all validations scenarios, only the ones from *feature4* were presented for the other feature sets had a similar pattern with 1 signifies low level, 2 represents medium level, and 3 denotes high level. Table 4-6 exhibit similar pattern that valence recognition struggled for medium class, while arousal recognition looked balance for all classes. On the other hand, recognizing low level of valence was easier than the other, and it was even the highest among all. Fixing problem in medium valence could improve the performance.

Table 4
Confusion matrix of feataure4 validated using subject-dependent

	Valence				Arousal		
	1	2	3		1	2	3
1	**0.907**	0.028	0.065	1	**0.770**	0.146	0.084
2	0.285	**0.485**	0.230	2	0.196	**0.793**	0.011
3	0.187	0.059	**0.754**	3	0.166	0.089	**0.745**

Table 5
Confusion matrix of *feataure4* validated using subject-independent

	Valence				Arousal		
	1	2	3		1	2	3
1	**0.925**	0.025	0.050	1	**0.781**	0.134	0.085
2	0.273	**0.485**	0.242	2	0.196	**0.793**	0.011
3	0.184	0.056	**0.760**	3	0.174	0.083	**0.743**

Table 6
Confusion matrix of *feataure4* validated using LOO validation

	Valence				Arousal		
	1	2	3		1	2	3
1	**0.910**	0.030	0.060	1	**0.756**	0.129	0.085
2	0.273	**0.500**	0.227	2	0.201	**0.788**	0.011
3	0.179	0.061	**0.760**	3	0.174	0.083	**0.743**

Further work is required to validate findings in this study using other databases. It is important to assess whether the proposed method truly offers superior features for emotion recognition by comparing the results utilizing other feature extraction method, e.g. nonlinear method, and experimenting with other classifiers. Since Neighborhood Components Analysis (NCA) showed powerful to enhance the emotion recognition using ECG-based features only [4], the same enhancement for the EDA-based feature is another interesting future works.

V. Conclusion

EDA-based emotion recognition using features from the cvxEDA method was presented. Standard and frequency-domain features as well as statistical distribution-based features were extracted for kNN classifier, validated using subject-dependent and subject-independent scenarios, and also LOO. In this study, the aim was to provide baseline recognitions for valence and arousal using the EDA-based feature only for the MAHNOB-HCI.

This study has identified that ratio of stimulated to relaxation stages as features from standard analysis and statistical distribution has no contribution after feature selection using sequential forward floating search. On the other hand, including this ratio as features from frequency-domain resulted in superior feature set, which outperformed the others.

Compare to some references [1], [3], [4], the proposed features used in this study offered better results for the 3-class problem, see highlighted results in Table 1 and 2. These are the second major findings in this study and will serve as baselines for future studies using the MAHNOB-HCI, especially for EDA-based features only. However, recognizing medium valence looked challenging, while low valence showed the easiest ones.

This study was limited by the absence of nonlinear features as Yang and Liu found that relationship between EDA signal and emotion is nonlinear [8]. Deeper studies using nonlinear features, e.g. Lanata et al. [6] proposed recurrent plot, deterministic chaos, and detrended fluctuation analysis, were left for future works. It would be interesting to assess the effects of NCA, which showed promising results in [4]. Applying the NCA can probably improve the negative finding on the confusion matrices, such that the medium valence recognitions are improved to increase the performances of the whole system.

Acknowledgment

This research was supported by the Finnish Cultural Foundation, Northern Ostrobothnia Regional Fund 2017.

References

[1] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A Multimodal Database for Affect Recognition and Implicit Tagging,” *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.

[2] H. Ferdinando, L. Ye, T. Seppänen, and E. Alasaarela, “Emotion Recognition by Heart Rate Variability,” *Aust. J. Basic Appl. Sci. Aust. J. Basic Appl. Sci*, vol. 8, no. 814, pp. 50–55, 2014.

[3] H. Ferdinando, T. Seppanen, and E. Alasaarela, “Comparing features from ECG pattern and HRV analysis for emotion recognition system,” in *2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2016, pp. 1–6.

[4] H. Ferdinando, T. Seppänen, and E. Alasaarela, “Enhancing Emotion Recognition from ECG Signals using Supervised Dimensionality Reduction,” in *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM*, 2017, pp. 112–118.

[5] S. D. Kreibig, “Autonomic nervous system activity in emotion: a review.,” *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, Jul. 2010.

[6] A. Lanata, G. Valenza, and E. P. Scilingo, “A novel EDA glove based on textile-integrated electrodes for affective computing,” *Med. Biol. Eng. Comput.*, vol. 50, no. 11, pp. 1163–1172, 2012.

[7] P. Ren, A. Barreto, Y. Gao, and M. Adjouadi, “Comparison of the use of pupil diameter and galvanic skin response signals for affective assessment of computer users,” *Biomed. Sci. Instrum.*, vol. 48, pp. 345–350, 2012.

[8] Z. Yang and G. Liu, “Emotion Recognition Based on Nonlinear Features of Skin Conductance Response,” *J. Inf. Comput. Sci.*, vol. 10, no. 12, pp. 3877–3887, Aug. 2013.

[9] A. Greco, G. Valenza, and E. P. Scilingo, *Advances in electrodermal activity processing with applications for mental health : from heuristic methods to convex optimization*. Cham, Switzerland: Springer

International Publishing AG, 2016.

[10] M. Benedek and C. Kaernbach, “A continuous measure of phasic electrodermal activity,” *J. Neurosci. Methods*, vol. 190, no. 1, pp. 80–91, 2010.

[11] M. Benedek and C. Kaernbach, “Decomposition of skin conductance data by means of nonnegative deconvolution,” *Psychophysiology*, vol. 47, no. 4, pp. 647–658, Mar. 2010.

[12] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. W. Picard, “Automatic Identification of Artifacts in Electrodermal Activity Data,” in *IEEE Engineering and Medicine in Biology Society*, 2015, pp. 1934–1937.

[13] A. Greco, G. Valenza, A. Lanata, E. Scilingo, and L. Citi, “cvxEDA: a Convex Optimization Approach to Electrodermal Activity Processing,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, 2016.

[14] H. Gunes and H. Hung, “Is automatic facial expression recognition of emotions coming to a dead end? The rise of the new kids on the block,” *Image Vis. Comput.*, vol. 55, pp. 6–8, 2016.

[15] P. Venables and M. Christie, “Electrodermal activity,” in *Techniques in psychophysiology*, I. Martins and P. Venables, Eds. New York: John Wiley & Sons Inc, 1980, pp. 3–67.