Open access • Journal Article • DOI:10.1109/JSEN.2016.2623677

Arousal and Valence Recognition of Affective Sounds Based on Electrodermal Activity

Alberto Greco, Gaetano Valenza, Luca Citi, Enzo Pasquale Scilingo

Institutions: University of Pisa, University of Essex

Published on: 01 Feb 2017 - IEEE Sensors Journal (IEEE)

Topics: Valence (psychology)

Arousal and Valence Recognition of Affective Sounds based on Electrodermal Activity

Alberto Greco, Gaetano Valenza, Luca Citi, and Enzo Pasquale Scilingo

***Abstract*—Physiological sensors and interfaces for mental healthcare are becoming of great interest in research and commercial fields. Specifically, biomedical sensors and related ad-hoc signal processing methods can be profitably used for supporting objective, psychological assessments. However, a simple system able to automatically classify the emotional state of a healthy subject is still missing. To overcome this important limitation, we here propose the use of convex optimization-based electrodermal activity (cvxEDA) framework and clustering algorithms to automatically discern arousal and valence levels induced by affective sound stimuli.**

EDA recordings were gathered from 25 healthy volunteers, using only one EDA sensor to be placed on fingers. Standardized stimuli were chosen from the International Affective Digitized Sound System database, and grouped into four different levels of arousal (i.e., the levels of emotional intensity) and two levels of valence (i.e., how unpleasant/pleasant a sound can be perceived). Experimental results demonstrated that our system is able to achieve a recognition accuracy of 77.33% on the arousal dimension, and 84% on the valence dimension.

***Index Terms*—Electrodermal Activity, Electrodermal Response, Sparse Representation, Convex Optimization, Emotion Recognition, Affective Digitized Sound System (IADS), K-NN Classifier.**

I. Introduction

The auditory channel is one of the most powerful means to induce and communicate emotions to people. It is easy to understand that speakers can elicit emotions through the voice tone conveying them to the hearers as well [1]. Music is another way to enhance orally expressed affective messages [2], [3]. But speech and music seem to be only a portion of the sounds that we hear. There are also non-musical and nonlinguistic sounds [4], which carry affective information in the audio environment around a listener [5].

Research on emotions indicates that affective stimuli can automatically elicit central and peripheral autonomic nervous system (ANS) reactions. Therefore ANS, among other measures, can offer reliable markers about emotional reactions [6]–[13]. In particular, several studies have been shown how ANS activity varies according to affective valence (i.e. pleasantness of the perception) and arousal (i.e. intensity of the perception), which are the two main dimensions of the circumplex model of affect (CMA) [14]

In the literature there are studies dealing with the relationship among physiological signals and pleasant and unpleasant sounds or music-induced emotions. PET and fMRI studies on emotions evoked by auditory stimuli have found that pleasant sounds lead to the activation of brain areas such as the orbitofrontal cortex and the anterior insula [15]. Sad music or unpleasant noises instead lead to the activation of regions involved in negative emotional states and anxiety-related, such as the hippocampus, the amygdala and the areas of the medial temporal lobe [16], [17]. All this brain areas are directly involved in the control of the human affective system. Other studies take into account EEG and ANS dynamics [18]–[23], also to automatically recognize four types of music-induced emotions [18].

In view of this, it is well known that there are specific peripheral activation patterns associated with the emotional valence and arousal of affective sounds. One of the most powerful peripheral measures of the ANS neural pathway is the Electrodermal Activity (EDA). This measure has advantages over other measures of the ANS, since EDA is under direct control of the sympathetic branch of the ANS. In fact, currently proposed heart rate variability measures, for instance, have not been proven to be effective in estimating sympathetic dynamics [24], [25]. Moreover, EDA has been shown to provide reliable quantification of autonomic expressions of emotions in the auditory domain [8]. In the literature, several studies have investigated on EDA dynamics in relationship to emotional changes elicited by music. A higher electrodermal reaction has been shown in response to emotionally valenced stimuli (pleasant or unpleasant) compared to neutral stimuli [26]. Concerning the arousal dimension, interesting results have shown that the electrodermal responses can be evoked and modulated by musical emotional arousal [27]–[29].
These findings have suggested the feasibility of developing a pattern recognition system to automatically classify arousal and valence levels of auditory stimuli. In the last decade, several studies have applied different classification procedures to feature-sets extracted from multi-modal recordings (including EDA) [21], [29], [30]. More specifically, EDA was acquired with other physiological signals, such as electrocardiogram, blood volume pulse, respiratory signal and electromyography. In these studies the feature selection stage showed that the most relevant features were computed from ECG signal. However, both the skin conductance mean level [31] and some statistical measures (e.g., mean, maximum, variance) computed from the filtered and derived EDA signal [21], [30]

The research leading to these results has received partial funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement n 601165 of the project WEARHAP (WEARABLE HAPTICS FOR HUMANS AND ROBOTS) .

A.G., G.V., and E.P.S. are with the department of Information Engineering and Research Center "E. Piaggio", School of Engineering, University of Pisa, Via G. Caruso 16 - 56122, Pisa, Italy.

L.C. is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K.

* Corresponding author email: alberto.greco@centropiaggio.unipi.it

contributed to increase the accuracy of the pattern recognition system. In addition, the EDA features looked more significant for arousal classification than valence [21].
Despite of these large number of results, only one of these studies has reported on an emotion recognition analysis using only the EDA signal with a very low accuracy (i.e., 56.5% for the arousal dimension and 50.5 % for the valence dimension) [30].
A crucial aspect for extracting suitable information from EDA analysis is the methodological approach, which may determine a large variability of results. Empirical evidences suggest that model-based methods (e.g., Pspm, CDA, cvxEDA ..) [32]–[34] have a greater validity compared to operational (i.e., heuristic and ad-hoc analyses) approaches [35]. In model-based approaches, models describe and estimate the underlying psychological process (i.e. ANS activity) that generates the observed data (i.e., EDA measurements). The model based analysis of EDA has fundamental advantages, such as a propensity to reduce the effects of measurement noise and the essential ability to improve the temporal resolution of inference in rapid event-related paradigms [33], [36]. In addition, statistical inference on the central autonomic network made by the EDA causal models is generally more powerful than statistical comparisons of observed data performed by operational approaches (e.g., standard peak detection). Furthermore, the model-based analysis gives explicit and quantitative description of assumptions that are, instead, implicit in operational approaches [35], [37], [38]. This allows a rigorous testing of those assumptions. In fact, operational approaches can be considered as implicit causal models that are not described in analytic terms. Model-based approaches explicate the implicit operational assumptions (e.g., the explicit modeling (and estimation) of observation noise) [35]. Among these models in the literature, cvxEDA algorithm [34] provides a window on the ANS activity, incorporating the physiological characteristic of the EDA by means of prior probabilities and constraints, without requiring pre- or post- processing steps and heuristic or ah-doc solutions.

The proposal of this computational study is to automatically recognize arousal and valence levels of the standardized affective acoustic stimuli, gathered from the International Affective Digitized Sound system (IADS) [39] dataset, using only the EDA analysis. Likewise the more used International Affective Picture System (IAPS) database [40], [41], IADS is a collection of sounds characterized in terms of valence and arousal [39]. Of note, the use of standardized stimuli (instead of, e.g., music) allows replicating studies in a more reliable fashion, making also easier the comparison of the results with future studies. IADS sounds have been already used in the literature (often jointly with IAPS picture) showing changes in ANS dynamic [10], [11] and a relationship with EDA variations [12], [13].

In order to perform this study we designed an experiment where twenty-five healthy subjects listened to IADS sounds. During the experimental sessions, EDA signal was continuously monitored. As will be explained in detail, EDA was processed using a recently validated approach called cvxEDA [34] in order to extract features strictly related to arousal and valence levels of the affective sounds and to perform an automatic pattern recognition.

In the next section, we will explain the rationale of the cvxEDA algorithm and the experimental protocol. Then, each processing stage will be described in detail, including feature extraction, statistical analysis and classification procedures. Section III will show the experimental results of both the statistical analysis and the pattern recognition analysis. Finally, results will be discussed in Section IV along with possible perspectives for future applications.

II. MATERIALS AND METHODS

The EDA is a general term that indicates variations in the electrical properties of the skin, which is specifically referred to as skin conductance (SC), due to the eccrine sweat gland activity. These glands are directly innervated by the sympathetic branch of the ANS and more specifically by the sudomotor nerve. Changes in the skin conductance of particular sites of the human body, such as fingers or palms of the hands, where the concentration of the eccrine glands is much higher than other body parts, can be related to both the psycho-physiological state of a person and the interaction with exogenous events [42]. The frequency band of the SC signal is in the range $0 - 2$ Hz [43] and can be split into a slowly-varying component, the so-called tonic component, and a superposed phasic component [9], [42], [44].

Tonic component corresponds to the baseline level of the SC signal (i.e., skin conductance level, SCL), and contains information about the subject's general psycho-physiological state, and his autonomic regulation [42]. The phasic component represents changes of the SC over a short time response window ($1 - 5$ s after stimulus onset [33]), in reaction to exogenous stimuli such as sounds, pictures, lights, smells. The skin conductance response (SCR) rises in a short time period but sometimes is not able to return to tonic levels before a subsequent response, resulting in an overlap of consecutive responses, with consequent loss of information. Sometimes, there are SCRs, which are significant from their background but not tied to a specific stimulus, typically referred to as Non-specific-SCR (NsSCR). They have the same characteristics as stimulus-related SCRs, but are considered tonic measures because they occur in the absence of external stimuli and in the absence of artifacts, e.g., movements [45].

In this section, we briefly report on the cvxEDA model, which is described in detail in [34]. Of note, this method is able to discern overlapping consecutive SCRs, likely to be present in case of an inter-stimulus interval shorter than the SCR recovery time.

A. EDA Processing using cvxEDA algorithm

CvxEDA proposed a representation of the SCRs as the output of a linear time-invariant system to a sparse non-negative driver signal. The model assumes that the observed SC (y) is the sum of the phasic activity (r), a slow tonic component (t), and an additive independent and identically distributed zero-average Gaussian noise term:

$$y = r + t + \epsilon. \tag{1}$$

Physiologically-plausible characteristics (temporal scale and smoothness) of the tonic input signal can be achieved by means of a cubic spline with equally-spaced knots every 10 s, an offset and a linear trend term:

$$t = B\ell + Cd, \tag{2}$$

where:

- B is a tall matrix whose columns are cubic B-spline basis functions,
- ℓ is the vector of spline coefficients,
- C is a $N \times 2$ matrix (where N is the length of the SC time series) with $C_{i,1} = 1$ and $C_{i,2} = i/N$,
- d is a 2×1 vector with the offset and slope coefficients for the linear trend.

The phasic component is the result of a convolution between the sudomotor nerve activity (SMNA), p, and an impulse response $h(t)$ shaped like a biexponential Bateman function [46]–[48]:

$$h(t) = (e^{-\frac{t}{\tau_1}} - e^{-\frac{t}{\tau_2}})\, u(t) \tag{3}$$

where τ_1 and τ_2 are, respectively, the slow and the fast time constants of the phasic curve shape, and $u(t)$ is the unitary step function.

Taking the Laplace transform of (3) and then its discrete-time approximation with sampling time δ (using a bilinear transformation), we obtain an autoregressive moving average (ARMA) model (see details in [34]) that can be represented in matrix form as

$$H = M^{-1}A, \tag{4}$$

where M and A are tridiagonal matrices with the MA and AR coefficients along the diagonals. Using an auxiliary variable q such that

$$q = A^{-1}\, p, \quad r = M\, q, \tag{5}$$

we write the final observation model as

$$y = Mq + B\ell + Cd + \epsilon. \tag{6}$$

Given the EDA model (6), the goal is to identify the maximum a posteriori (MAP) neural driver SMNA (p) and tonic component (t) parametrized by $[q, \ell, d]$, for the measured SC signal (y). CvxEDA rewrites the MAP problem as a constrained minimization QP convex problem (see details in [34]):

$$\begin{aligned} &\text{minimize } \frac{1}{2}\left\|Mq + B\ell + Cd - y\right\|_2^2 + \alpha\delta\left\|Aq\right\|_1 + \frac{\gamma}{2}\left\|\ell\right\|_2^2 \\ &\text{subj. to } Aq \geq 0. \end{aligned} \tag{7}$$

After some matrix algebra, this optimization problem can be re-written in the standard QP form and solved efficiently using one of the many sparse-QP solvers available. After finding the optimal $[q, \ell, d]$, the tonic component t can be derived from (2) while the sudomotor nerve activity driving the phasic component can be easily found as $p = Aq$.

The objective function (7) to be minimized is a quadratic measure of misfit or prediction error between the observed data and the values predicted by the model. Moreover, the prior knowledge about the spiking sparse nature and nonnegativity of the SMNA (p) and the smoothness of the tonic component are accounted for by the regularizing terms and the constraint.

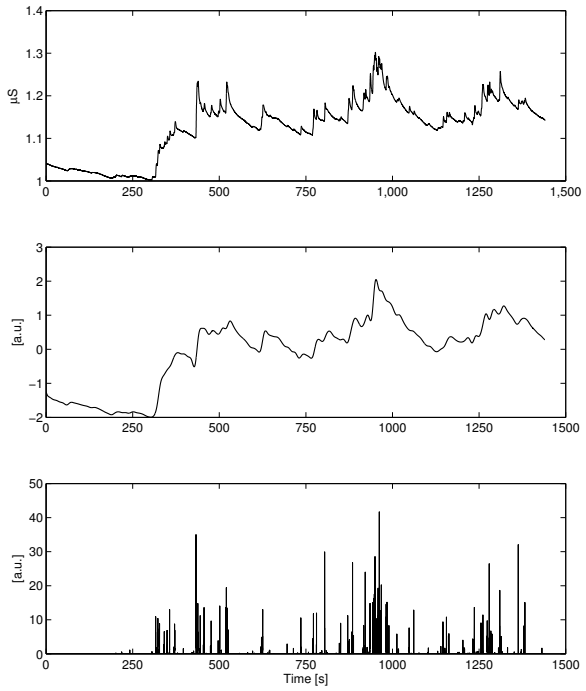

Fig. 1. Application of the cvxEDA decomposition procedure to the SC signal recorded during the affective sound stimulation for a representative subject. (Top) Raw SC signal, (Middle) Estimated slow tonic component t, (Bottom) Estimated sparse phasic driver component p.

The strength of the penalty is regulated by α and γ terms. A sparser estimate is yielded by large values of α. Concerning γ, higher values mean a stronger penalization of ℓ, i.e. a smoother tonic curve.

Of note, an example application of the cvxEDA algorithm to an EDA signal acquired during affective sound stimulations is shown in Figure 1. It is worthwhile noting the smoothed nature of the tonic component as well as the sparse representation of the phasic component, which represent the neural sudomotor response to the external stimulation.

B. Subjects Recruitment, Experimental protocol and Acquisition set-up

In this study, we enrolled twenty-five healthy subjects (aged 25–35) who voluntarily underwent the experiment. All subjects signed a written informed consent prior to taking part in the study, and the protocol was approved by the local Ethics Committee. Moreover, according to the self-report questionnaires, none of them had a history of injury of the auditory canal or partial or full incapability of hearing. Moreover, none of them suffered from any mental or chronic disease.

In order to avoid any bias, subjects were informed about the purpose of the study and the kind of emotional stimulation

which was going to be administered, but they were not aware of the level of intensity (i.e. arousal) and pleasantness (i.e. valence) they would have been listened to. During the experiment, participants were seated in a comfortable chair in a controlled environment while listening to the IADS sounds. Each subject was left alone in the room where the experiment took place for the whole duration (29 minutes). The acoustic stimulation was performed by using headphones while the subject's eyes were closed, to avoid any kind of visual interference.

The experimental protocol consisted of 8 sessions: after an initial resting session of 5 minutes, three arousal sessions alternated with neutral sessions (see Figure 2). Each arousal and neutral session was different from the others in regard to the arousal level (labeled as N (neutral), L (low), M (medium) and H (high)). We selected four arousal ranges that were not overlapped. Such levels were set according to the IADS scores reported in table I. Within each arousing session, the acoustic stimuli were selected to have both negative and positive valence. Each neutral session lasted 1 minute and 28 seconds, while the three arousal sessions had a duration of 3 minutes and 40 seconds, 4 minutes, and 5 minutes and 20 seconds, respectively. The different duration of each arousal session is due to the different length of acoustic stimuli having the same range of positive and negative valence.

During the elicitation, the EDA was continuously acquired, by means of a dedicate hardware module of the BIOPAC MP150 acquisition system. The two electrodes were placed on the distal phalanges of the index and middle fingers.

TABLE I
RATING OF IADS SOUNDS USED IN THIS WORK

Session	N. of Sounds	Valence Rating	Valence Range	Arousal Rating	Arousal Range
Neutral	8	5.915±0.68	4.34÷6.44	3.47±0.175	2.88÷3.93
Arousal 1	19	/	2.46÷7.78	5.42±0.22	5.00÷5.89
Arousal 2	26	/	2.04÷7.90	6.48±0.25	6.00÷6.99
Arousal 3	20	/	1.57÷7.67	7.32±0.22	7.03 ÷8.16

Ratings are expressed as median and its absolute deviation.

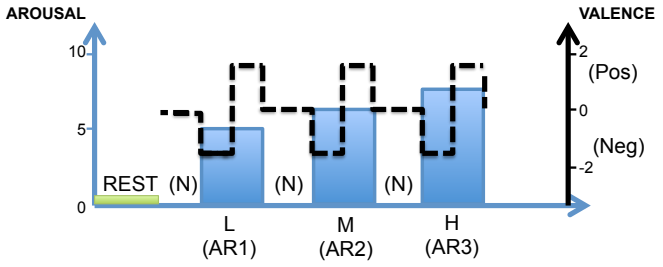

Fig. 2. Timeline of the experimental protocol in terms of arousal and valence levels. The vertical axes relate to the arousal (blue) and valence (black) IADS score, whereas the horizontal axis relates to the time. The neutral sessions alternate with the arousal ones, which are marked with blue staircases. The three arousal sessions have increasing intensity of activation (i.e., L (low), M (medium) and H (heigh)) over time. The dotted black line indicates the valence levels within an arousing session (i.e., Pos (positive) and Neg (negative)). The neutral sessions are characterized by lowest arousal and medium valence scores. Green line relates to the resting state.

C. Signal Processing

The convex-optimization-based EDA model (cvxEDA) described in II-A was applied to each participant's SC signal. Before cvxEDA application, each time series was normalized by means of a Z-score transformation in order to standardize the dataset and increase the velocity of the optimization procedure.

Concerning the cvxEDA parameters, the values $\alpha = 0.008$ and $\gamma = 0.01$ were employed for all subjects. Moreover, we assumed a canonical IRF for all subjects with $\tau_1 = 0.7$ and $\tau_2 = 3$. These values were chosen through an exploratory analysis on simulated data and real signals from a different dataset. In these exploratory studies, rather than assuming a canonical IRF for all subjects, we assumed $\tau_1 = 0.7$ and adapted the parameter τ_2 on a per-subject basis. After fitting cvxEDA models for several values of $\tau_2 \in [2.0, 4.0]$, we chose the one minimizing the L_0-"norm", i.e. leading to the sparsest solution [9], [34], [49].

It is worthwhile noting that the convex optimization approach does not need any pre-processing step. In fact, unlike other approaches in the literature, this model relies exclusively on the presence and definition of the priors (see [34]) to impose physiologically sound constraints. All the measurement and modelling errors are included in the estimated noise term ϵ without any need of pre-processing step.

Statistical analysis and pattern recognition methodologies were used in this study to automatically recognize the autonomic response of the subjects to the emotional stimulus, and to associate the values of the EDA parameters with the kind of elicitation as expressed in terms of arousal and valence.

D. Feature Extraction and Statistical analysis

The autonomic neural activity was quantified computing several features from both the tonic and phasic components that are outputs of the cvxEDA. The mean value of the tonic component was calculated in order to estimate the general psychophysiological status of the subjects. The characterization of the stimulus response was achieved from the sparse phasic signal. Specifically, we calculated the number of the peaks, their maximum amplitude and the area under curve within a time response window of 5 sec after each stimulus onset (see Table II).

TABLE II
LIST OF FEATURES EXTRACTED FROM EDA PHASIC AND TONIC COMPONENTS.

Feature	Description
Npeak	number of significant SMNA SCR wrw
AUC	Area under curve of SMNA signal wrw (μSs)
peak	maximum amplitude of significant peaks of SMNA signal wrw[1](μS)
stdphasic	Standard deviation of SMNA signal wrw (μS)
MeanTonic	Mean value of the tonic component within session windows (μS)

wrw= within response window (i.e., 5 sec after stimulus)

For each feature, the two levels of valence, i.e., positive and negative, and the three arousal levels (Low, Medium and High) were statistically compared. Each statistical comparison

is performed using the nonparametric Wilcoxon tests for paired samples, given the non-gaussianity of data ($p < 0.05$ from Shapiro-Wilk test with null hypothesis of having a Gaussian sample). The multiple pairwise comparisons among the three arousal levels was corrected with the Bonferroni's method, multiplying the p-values by the number of comparisons, i.e., 3. Furthermore, the multivariate pattern analysis was performed in order to classify different levels of arousal and two levels of valence.

Furthermore, a multivariate pattern analysis was performed in order to classify the three levels of arousal and the two levels of valence, as specified in the next paragraph.

Of note, concerning the arousal analysis, we considered three different datasets:

- Complete Dataset (CD): the features were extracted from each digital sound of each whole arousal session and then the values were averaged;
- Positive Dataset (PD): the features were extracted from each digital sound of the arousal session that have a positive valence, and then the values were averaged;
- Negative Dataset (ND): the features were extracted from each digital sound of the arousal session that have a negative valence, and then the values were averaged;

The comparison among the three arousal levels were performed for each dataset described above, in order to study eventual effect of the valence in the arousal discrimination.

E. Classification Procedure

The supervised classification of input feature data in the pattern recognition method was implemented following a Leave-One-Subject-Out procedure (LOSO) and using a K-NN Classifier. We applied a feature selection on each training set comprised of $(N-1)$ subjects (where N is the total number of participants) to recognize the emotional responses of the subject N_{th}. This procedure was iterated N times. In consequence of the way the experiment was designed, a paired dataset had to be considered. Therefore, the matrix of the input data was transformed in a within-subject rank matrix (based on the Friedman statistical test). More specifically, for each feature the examples relative to a single subject were transformed in ranks. For example, regarding the arousal classification, for each subject, each of the m features was represented through a vector of three values corresponding to the three arousal levels. This vector was ranked from 1 to 3. The three ranked-values were a coordinate of four points in the m-dimensional feature-space

Two separate classification algorithms were implemented for the arousal and valence recognition. For each of the two classification algorithms, the dimension of the feature space corresponded to the number of selected EDA parameters. The final feature set was achieved identifying the combination of features and classifier parameters that performed the highest recognition accuracy in the training set. Specifically, on each $(N-1)$ training set we applied a nested-LOSO, which iteratively selected a nested-traning set of $(N-2)$ subjects to find out the most accurate combination of features and number of neighbors. For the arousal recognition problem, the number of samples (i.e., rows of the dataset) in such a space was related to the number of subjects multiplied by the number of arousal levels (i.e., 25x3=75) to be classified. Concerning the valence classification, instead, we had a negative valence example and a positive one for each arousal session for each subject (i.e., 25x3x2=150). Finally, the ranked feature set was used as input of a K-NN Classifier [50], [51], which was validated through the Leave-One-Subject-Out (LOSO) procedure [52]. A block diagram of the proposed recognition system is illustrated in Figure 3. The classification results were expressed as recognition accuracy in form of confusion matrix [53]. A generic element c_{ij} of the confusion matrix indicated the percentage of how many times the feature set belonging to the class i, was recognized as belonging to the class j. This means that a higher average of the values on the matrix diagonal corresponds to a better degree of classification.

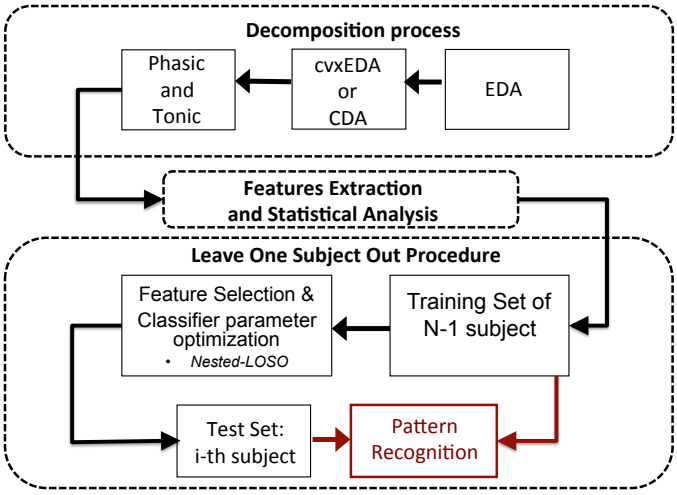

Fig. 3. Overall block scheme of the proposed arousal and valence recognition system. The EDA is processed in order to extract the phasic and tonic components using the cvxEDA algorithm. According to the protocol timeline, several features are extracted and, then, statistically compared. The KNN classifier is engaged to perform the pattern recognition by adopting a leave-one-subject-out procedure.

III. Experimental Results

In this section we will describe the results coming from both the statistical and the pattern recognition analysis applied separately to the following datasets: Arousal Dataset (CD, PD, ND), and Valence Dataset (see Tables III, V, VI, and IV).

1) Statistical analysis results: All the features extracted from the phasic component did not show any significant difference among the three levels of arousal (as example, Fig. 4 plotted a statistical description of the comparison among the maximum phasic peaks). The same results were confirmed even when the dataset were divided in the PD and ND.

Instead, the mean tonic value statistically discriminates the three arousal and the two valence levels, showing a monotonic trend, although inverse (Figure 5). Moreover, considering both the ND and the PD the trend of the tonic feature was the same as above. In addition, the tonic value showed a significant statistical difference among the three arousal sessions.

TABLE III
THE FIRST THREE COLUMNS SHOW THE MEAN ± STANDARD DEVIATION VALUE OF EACH CVXEDA FEATURES AMONG THE SUBJECTS CONSIDERING THE COMPLETE DATASET (CD). THE LAST THREE COLUMNS SHOW THE P-VALUES OF THE PAIRWISE STATISTICAL COMPARISONS PERFORMED BY MEANS OF WILCOXON TEST WITH BONFERRONI CORRECTION AMONG THE THREE AROUSAL (I.E., AR1, AR2, AR3) LEVELS.

Feature of CD	AR1	AR2	AR3	p-value		
				AR1-AR2	AR1-AR3	AR2-AR3
Npeak	2.99 ± 4.32	2.85 ± 5.095	3.26 ± 3.61	> 0.05	> 0.05	> 0.05
Peak	63.94 ± 255.094	29.053 ± 94.31	106.45 ± 387.087	> 0.05	> 0.05	> 0.05
AUC	411.94 ± 1719.27	424.44 ± 1825.71	571.46 ± 2090.83	> 0.05	> 0.05	> 0.05
stdphasic	4.67 ± 17.24	2.87± 9.25	10.38 ± 37.90	> 0.05	> 0.05	> 0.05
MeanTonic	-2.47 ± 13.31	-2.84 ± 14.56	-3.83 ± 16.53	**0.017**	**5.82 10^{-04}**	**0.0036**

TABLE IV
THE FIRST TWO COLUMNS SHOW THE MEAN ± STANDARD DEVIATION VALUE OF EACH CVXEDA FEATURES AMONG THE SUBJECTS CONSIDERING THE COMPLETE DATASET (CD). THE LAST COLUMN SHOW THE P-VALUES OF THE STATISTICAL COMPARISONS PERFORMED BY MEANS OF WILCOXON TEST WITH BONFERRONI CORRECTION AMONG THE TWO VALENCE (I.E.,POS AND NEG) LEVELS.

Feature of CD	Pos	Neg	p-value
Npeak	3.12 ± 4.22	3.28± 4.55	> 0.05
Peak	59.49 ± 233.81	52.67±240.27	> 0.05
AUC	474.23± 1836.25	429.35 ± 1701.42	> 0.05
stdphasic	5.71 ± 22.40	4.73 ± 20.54	> 0.05
MeanTonic	-2.98 ± 14.31	-2.64± 13.25	**6.64 10^{-04}**

TABLE V
THE FIRST THREE COLUMNS SHOW THE MEAN ± STANDARD DEVIATION VALUE OF EACH CVXEDA FEATURES AMONG THE SUBJECTS, CONSIDERING THE NEGATIVE DATASET (ND). THE LAST THREE COLUMNS SHOW THE P-VALUES OF THE PAIRWISE STATISTICAL COMPARISONS PERFORMED BY MEANS OF WILCOXON TEST WITH BONFERRONI CORRECTION AMONG THE THREE AROUSAL (I.E., AR1, AR2, AR3) LEVELS.

Featureof ND	AR1	AR2	AR3	p-value		
				AR1-AR2	AR1-AR3	AR2-AR3
Npeak	2.97 ± 4.58	2.96 ± 5.33	3.22 ± 3.95	> 0.05	> 0.05	> 0.05
Peak	64.51 ± 255.82	17.31 ± 47.49	96.68 ± 398.87	> 0.05	> 0.05	> 0.05
AUC	403.99 ± 1660.72	412.10 ± 1830.50	509.46 ± 1974.08	> 0.05	> 0.05	> 0.05
stdphasic	4.64 ± 16.78	1.80 ± 5.21	9.09 ± 36.30	> 0.05	> 0.05	> 0.05
MeanTonic	-2.43 ± 13.01	-2.69 ± 14.31	-3.42 ± 15.60	**0.033**	**0.0015**	**0.0027**

TABLE VI
THE FIRST THREE COLUMNS SHOW THE MEAN ± STANDARD DEVIATION VALUE OF EACH CVXEDA FEATURES AMONG THE SUBJECTS, CONSIDERING THE POSITIVE DATASET (PD). THE LAST THREE COLUMNS SHOW THE P-VALUES OF THE PAIRWISE STATISTICAL COMPARISONS PERFORMED BY MEANS OF WILCOXON TEST WITH BONFERRONI CORRECTION AMONG THE THREE AROUSAL (I.E., AR1, AR2, AR3) LEVELS.

Feature of PD	AR1	AR2	AR3	p-value		
				AR1-AR2	AR1-AR3	AR2-AR3
Npeak	3.01 ± 4.15	2.85 ± 4.90	3.32 ± 3.59	> 0.05	> 0.05	> 0.05
Peak	63.37 ± 254.73	41.75 ± 149.26	104.73 ± 359.77	> 0.05	> 0.05	> 0.05
AUC	419.90 ± 1779.27	440.49 ± 1835.30	637.21 ± 2254.75	> 0.05	> 0.05	> 0.05
stdphasic	4.70 ± 17.77	4.045 ± 14.16	11.037 ± 37.99	> 0.05	> 0.05	> 0.05
MeanTonic	-2.52 ± 13.61	-2.98 ± 14.80	-4.23 ± 17.51	**7.99 10^{-04}**	**3.032 10^{-04}**	**0.0136**

2) Pattern recognition analysis results: Experimental results of the classification procedure are shown in the form of confusion matrix and reported in Table VII. The principal diagonal represents the percentage of the successful recognition of each class.

Concerning the arousal levels, the multivariate pattern recognition analysis showed a high accuracy (80.00%) considering both the arousal sessions as a whole (Table VII), only the positive valence stimuli (78.67%) (Table IX), and only the negative valence stimuli (77.33%) (Table VIII). Although the average accuracy along the diagonal was the same, the PD yielded a better discrimination of the first arousal stimulation, whereas the ND allowed a better recognition of the higher arousal classes. Moreover, the feature selection procedure pointed out that for the PD the maximum recognition accuracy was obtained using a combination of phasic AUC and peak features and meanTonic, instead for the ND the features selected were meanTonic and stdphasic.

Concerning the valence levels, the features extracted from the cvxEDA model showed a good ability in characterization of the autonomic activity. In fact, the classification procedure was able to discern the pleasant and unpleasant sounds with an accuracy of 84% (Table IX).

3) Feature selection results: The feature selection stage highlighted the most informative features (i.e., those that achieved the best accuracy in the training phase). Table XI shows for each dataset the informative power of each feature. With "informative power" of a specific feature we mean the

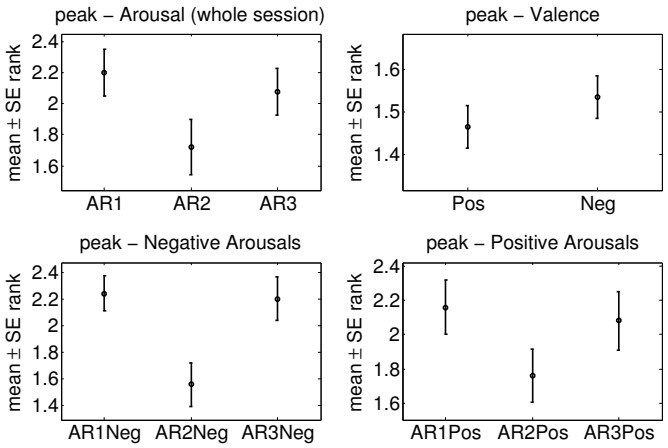

Fig. 4. Within-subject ranks of the maximum of the phasic component for the three arousal levels (i.e., AR1, AR2, AR3) (top-left), the two valence levels (i.e., Pos and Neg) (top-right), the three negative arousal levels (i.e., AR1Neg, AR2Neg, AR3Neg) (bottom-left), the three positive arousal levels (i.e., AR1Pos, AR2Pos, AR3Pos) (bottom-right). The dots mark the across-subject average rank for each level while the wiskers indicate the standard error. All arousal and valence comparisons were statistical compared by the Bonferroni-corrected Wilcoxon test that did not show significant differences.

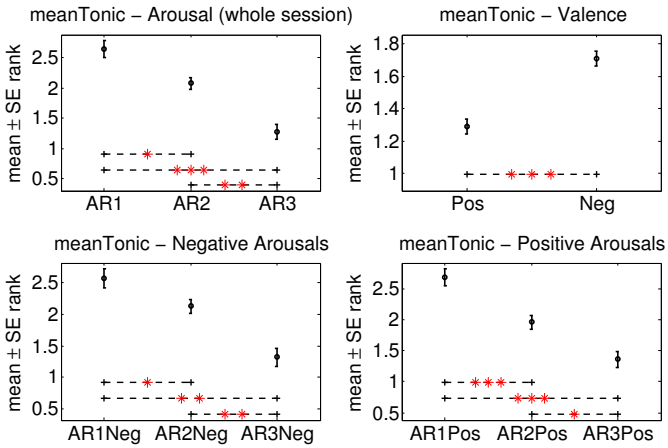

Fig. 5. Within-subject ranks of the mean tonic value for the three arousal levels (i.e., AR1, AR2, AR3) (top-left), the two valence levels (i.e., Pos and Neg) (top-right), the three negative arousal levels (i.e., AR1Neg, AR2Neg, AR3Neg) (bottom-left), the three positive arousal levels (i.e., AR1Pos, AR2Pos, AR3Pos) (bottom-right). The dots mark the across-subject average rank for each level while the wiskers indicate the standard error. Wilcoxon test and post-hoc Bonferroni-corrected pair-wise comparisons of the peak amplitudes found significant differences in the cases indicated by asterisks ($*: p < 0.05$; $**: p < 0.01$; $***: p < 0.001$).

TABLE VII
CONFUSION MATRIX FOR AROUSAL LEVEL (AR1, AR2, AR3) RECOGNITION BASED ON THE COMPLETE DATASET OF FEATURES EXTRACTED FROM THE CVXEDA MODEL OUTPUTS

	AR1	**AR2**	**AR3**
AR1	**72%**	8%	8%
AR2	12%	**88%**	12%
AR3	16%	4%	**80%**

percentage of iterations of the LOSO procedure where that feature has been selected over the total number N of iterations. The tonic mean level was always selected in each LOSO iteration of each input-dataset and resulted to be the most informative feature. Concerning the arousal classification, the AUC of the SMNA signal was highly informative (84%) when the input-dataset included the complete set of sounds (i.e., CD) and only the positive ones (i.e., PD). Considering the negative dataset, together with the mean tonic value, the other most selected features were the stdphasic (informative power of 92%) and the number of SMNA peaks (informative power of 20%).

On the other hand, concerning the valence classification problem, for most of LOSO-iterations, the feature set was complete with no selection.

TABLE VIII
CONFUSION MATRIX FOR NEGATIVE (N) AROUSAL LEVEL (AR1, AR2, AR3) RECOGNITION BASED ON THE NEGATIVE DATASET (NG) OF FEATURES EXTRACTED FROM THE CVXEDA MODEL OUTPUTS

	AR1 N	**AR2 N**	**AR3 N**
AR1 N	**72%**	12%	8%
AR2 N	16%	**80%**	12%
AR3 N	12%	8%	**80%**

TABLE IX
CONFUSION MATRIX FOR POSITIVE (P) AROUSAL LEVEL (AR1, AR2, AR3) RECOGNITION BASED ON THE POSITIVE DATASET (PD) OF FEATURES EXTRACTED FROM THE CVXEDA MODEL OUTPUTS.

	AR1 P	**AR2 P**	**AR3 P**
AR1 P	**80%**	16%	8%
AR2 P	8%	**84%**	20%
AR3 P	12%	0%	**72%**

TABLE X
CONFUSION MATRIX FOR VALENCE LEVEL RECOGNITION BASED ON THE COMPLETE DATASET OF CVXEDA FEATURES

	Positive valence	**Negative valence**
Positive valence	**84%**	16%
Negative valence	16%	**84%**

TABLE XI
INFORMATIVE POWER OF EACH FEATURE FOR EACH INPUT-DATASET

Classification problem	Feature	Informative power
Arousal	Npeak	8%
	AUC	84%
	peak	4%
	stdphasic	4%
	MeanTonic	100%
Negative Arousal	Npeak	16%
	AUC	84%
	peak	0%
	stdphasic	0%
	MeanTonic	100%
Positive Arousal	Npeak	20%
	AUC	0%
	peak	0%
	stdphasic	92%
	MeanTonic	100%
Valence	Npeak	72%
	AUC	84%
	peak	88%
	stdphasic	96%
	MeanTonic	100%

IV. Discussion and Conclusions

This study aimed at exploring the possibility of obtaining an automatic recognition of arousal and valence levels in twenty-five healthy volunteers which were stimulated by sets of affective sounds gathered from the IADS database. The automatic recognition procedure was performed on features coming from the analysis of the electrodermal dynamics exclusively. Stimuli were expressed in terms of arousal levels according to the circumplex model of affect, whereas the EDA was analyzed through a new model previously validated in [34]. The experimental protocol consisted in three levels of arousal stimuli, each of which containing pleasant and unpleasant sounds.

The statistical comparison among the three arousal groups of stimuli and the two valence levels showed no significant differences in the phasic feature set. On the contrary, the mean tonic value was significantly different among the three arousal levels and the two valence levels. Therefore, we were not able to statistically discriminate the phasic response to different stimuli, but the long sessions of stimulation were able to induce a significant change in the general psycho-physiological state of the subjects highly correlated with the tonic component and highlighted by the tonic features.

The classification procedure was applied to four different datasets, in order to discriminate the two valence levels and the three arousal levels taking into account only the positive sounds, only the negative sounds, and both the positive and negative stimuli without discriminations. Results thanks to the use of the cvxEDA outputs showed a recognition accuracy of 80.00% on the arousal dimension. Regarding only the positive and negative sounds the arousal recognition results confirmed a high discrimination power ($>$77%), pointing out how the arousal classification was minimally affected by valence dimension. Moreover, very good performance was obtained also in the valence classification. The-2 class problem achieved 84% of accuracy on all classes. Likewise the statistical analysis, the feature selection in the classification procedure suggested that ANS measures related to the tonic component are the most effective to discriminate and recognize emotional states induced by affective sounds. Of note, the phasic features contributed mostly to the valence classification rather than arousal recognition.

To our knowledge, this study shows for the first time the use of a sympathetic nervous system measures, which can be profitably acquired through a simple EDA sensor placed on fingers, to automatically recognize emotional states induced by affective sounds taken from IADS collection. It is also worthwhile noting that we were able to discern emotional stimuli by using features from our cvxEDA model, confirming the applicability of a sparse representation for emotion recognition.

Of note, cvxEDA applicability and good performance were already shown in [49] and [34] during affective touch and affective visual stimulations respectively. In addition, emotion recognition, through EDA features only, was already performed using visual stimuli, with and without cvxEDA application [54], [55]. To our knowledge, the use of EDA exclusively to recognize emotions induced by standardized auditory stimuli is the major novelty of this study. It is also worthwhile noting that we were able to discern emotional stimuli by using features from our cvxEDA model, confirming the applicability of a sparse representation for emotion recognition.

Moreover, outcomes of this research might have an impact in clinical research fields such as psychology or neurology. The use of visual emotional elicitations [56], in fact, could not be applied to patients with visual impairment, or patients with disorder of consciousness. Knowledge on sympathetic dynamics, as estimated through cvxEDA models, in healthy subjects might provide useful biomarkers to support clinical decisions.

Future work will investigate a parameter optimization procedure for the classifiers, as well as will focus on a rigorous comparison with other automatic classification methods, to be benchmarked within a larger sample size. Moreover, we plan to study the applicability of cvxEDA model on signals acquired from mobile wearable sensors (where noise filtering is a crucial step), and to port the cvxEDA algorithm to these wearable systems.

References

[1] C. Gobl and A. Ni, "The role of voice quality in communicating emotion, mood and attitude," *Speech communication*, vol. 40, no. 1, pp. 189–212, 2003.

[2] A. Goldstein, "Thrills in response to music and other stimuli," *Physiological Psychology*, vol. 8, no. 1, pp. 126–129, 1980.

[3] S. Johna, "Music structure and emotional response: Some empirical findings," *Psychology of music*, vol. 991, no. 9, p. L120, 1991.

[4] M. Marcell, M. Malatanos, C. Leahy, and C. Comeaux, "Identifying, rating, and remembering environmental sound events," *Behavior research methods*, vol. 39, no. 3, pp. 561–569, 2007.

[5] W. W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological psychology*, vol. 5, no. 1, pp. 1–29, 1993.

[6] R. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *Affective Computing, IEEE Transactions on*, vol. 1, no. 1, pp. 18–37, 2010.

[7] M. Nardelli, G. Valenza, A. Greco, A. Lanata, and E. Scilingo, "Recognizing emotions induced by affective sounds through heart rate variability," *Affective Computing, IEEE Transactions of*, vol. 6, no. 4, pp. 385–394, 2015.

[8] P. Lang, M. Bradley, and B. Cuthbert, "Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology," *Biological Psychiatry*, vol. 44, no. 12, pp. 1248–1263, 1998.

[9] A. Greco, A. Lanata, G. Valenza, E. P. Scilingo, and L. Citi, "Electrodermal activity processing: A convex optimization approach," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 2290–2293.

[10] J. Anttonen and V. Surakka, "Emotions and heart rate while sitting on a chair," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2005, pp. 491–499.

[11] L. Ivonin, H.-M. Chang, W. Chen, and M. Rauterberg, "Unconscious emotions: quantifying and logging something we are not aware of," *Personal and ubiquitous computing*, vol. 17, no. 4, pp. 663–673, 2013.

[12] E. Verona, C. J. Patrick, J. J. Curtin, M. M. Bradley, and P. J. Lang, "Psychopathy and physiological response to emotionally evocative sounds." *Journal of abnormal psychology*, vol. 113, no. 1, p. 99, 2004.

[13] A. Hariharan and M. T. P. Adam, "Blended emotion detection for decision support."

[14] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[15] A. J. Blood and R. J. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion," *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, pp. 11 818–11 823, 2001.

[16] S. Koelsch, T. Fritz, K. Müller, A. D. Friederici *et al.*, "Investigating emotion with music: an fMRI study," *Human brain mapping*, vol. 27, no. 3, pp. 239–250, 2006.

[17] S. Anders, F. Eippert, N. Weiskopf, and R. Veit, "The human amygdala is sensitive to the valence of pictures and sounds irrespective of arousal: an fMRI study," *Social cognitive and affective neuroscience*, vol. 3, no. 3, pp. 233–243, 2008.

[18] Y. Lin, C. Wang, T. Jung, T. Wu, S. Jeng, J. Duann, and J. Chen, "Eeg-based emotion recognition in music listening," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 7, pp. 1798–1806, 2010.

[19] Y.-H. Yang and H. H. Chen, *Music Emotion Recognition*. CRC Press, 2011.

[20] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen, "A regression approach to music emotion recognition," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 2, pp. 448–457, 2008.

[21] J. Kim and E. Andre, "Emotion recognition based on physiological changes in music listening," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 12, pp. 2067–2083, 2008.

[22] A. L. Roque, V. E. Valenti, H. L. Guida, M. F. Campos, A. Knap, L. C. M. Vanderlei, L. L. Ferreira, C. Ferreira, and L. C. de Abreu, "The effects of auditory stimulation with music on heart rate variability in healthy women," *Clinics*, vol. 68, no. 7, pp. 960–967, 2013.

[23] M. Orini, R. Bailon, R. Enk, S. Koelsch, L. Mainardi, and P. Laguna, "A method for continuously assessing the autonomic response to music-induced emotions through HRV analysis," *Medical & biological engineering & computing*, vol. 48, no. 5, pp. 423–433, 2010.

[24] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[25] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon, "Power spectral density analysis of electrodermal activity for sympathetic function assessment," *Annals of biomedical engineering*, pp. 1–12, 2016.

[26] M. Bradley and P. Lang, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 2, pp. 204–215, 2000.

[27] M. D. van der Zwaag, J. H. Janssen, and J. H. Westerink, "Directing physiology and mood through music: Validation of an affective music player," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 57–68, 2013.

[28] S. Khalfa, P. Isabelle, B. Jean-Pierre, and R. Manon, "Event-related skin conductance responses to musical emotions in humans," *Neuroscience letters*, vol. 328, no. 2, pp. 145–149, 2002.

[29] K. Cosic, S. Popovic, D. Kukolja, B. Dropuljic, D. Ivanec, and M. Tonkovic, "Multimodal analysis of startle type responses," *Computer Methods and Programs in Biomedicine*, 2015.

[30] X. Niu, L. Chen, H. Xie, Q. Chen, and H. Li, "Emotion pattern recognition using physiological signals," *Sensors and Transducers*, vol. 172, no. 6, pp. 147–156, 2014.

[31] K. Stephenson, E. Quintin, and M. South, "Age-related differences in response to music-evoked emotion among children and adolescents with autism spectrum disorders," *Journal of Autism and Developmental Disorders*, vol. 46, no. 4, pp. 1142–1151, 2016.

[32] D. R. Bach, "A head-to-head comparison of SCRalyze and Ledalab, two model-based methods for skin conductance analysis," *Biological Psychology*, vol. 103, pp. 63–68, 2014.

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

[34] A. Greco, G. Valenza, A. Lanata, E. Scilingo, and L. Citi, "cvxEDA: a convex optimization approach to electrodermal activity processing," *Biomedical Engineering, IEEE Transactions on*, 2015.

[35] D. R. Bach and K. J. Friston, "Model-based analysis of skin conductance responses: Towards causal models in psychophysiology," *Psychophysiology*, vol. 50, no. 1, pp. 15–22, 2013.

[36] R. J. Barry, S. Feldmann, E. Gordon, K. I. Cocker, and C. Rennie, "Elicitation and habituation of the electrodermal orienting response in a short interstimulus interval paradigm," *International Journal of Psychophysiology*, vol. 15, no. 3, pp. 247–253, 1993.

[37] D. R. Bach, K. J. Friston, and R. J. Dolan, "An improved algorithm for model-based analysis of evoked skin conductance responses," *Biological Psychology*, vol. 94, no. 3, pp. 490–497, 2013.

[38] D. R. Bach, G. Flandin, K. J. Friston, and R. J. Dolan, "Time-series analysis for rapid event-related skin conductance responses," *Journal of neuroscience methods*, vol. 184, no. 2, pp. 224–234, 2009.

[39] M. M. Bradley and P. J. Lang, "The international affective digitized sounds (iads): Affective ratings of sounds and instruction manual," *University of Florida, Gainesville, FL, Tech. Rep. B-3*, 2007.

[40] P. J. Lang, M. M. Bradley, B. N. Cuthbert *et al.*, *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. NIMH, Center for the Study of Emotion & Attention, 2005.

[41] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (iaps): Technical manual and affective ratings," 1999.

[42] W. Boucsein, *Electrodermal activity*, 2nd ed. Springer Science & Business Media, 2012.

[43] A. Ishchenko and P. Shev'ev, "Automated complex for multiparameter analysis of the galvanic skin response signal," *Biomedical Engineering*, vol. 23, no. 3, pp. 113–117, 1989.

[44] A. Greco, G. Valenza, A. Lanata, G. Rota, and E. Scilingo, "Electrodermal activity in bipolar patients during affective elicitation," *IEEE Journal of Biomedical and Health Informatics*, 2014.

[45] W. T. Roth, M. E. Dawson, and D. L. Filion, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 49, pp. 1017–1034, 2012.

[46] E. Garrett, "The bateman function revisited: a critical reevaluation of the quantitative expressions to characterize concentrations in the one compartment body model as a function of time with first-order invasion and first-order elimination," *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 22, no. 2, pp. 103–128, 1994.

[47] D. Alexander, C. Trengove, P. Johnston, T. Cooper, J. August, and E. Gordon, "Separating individual skin conductance responses in a short interstimulus-interval paradigm," *Journal of neuroscience methods*, vol. 146, no. 1, pp. 116–123, 2005.

[48] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *Journal of neuroscience methods*, vol. 190, no. 1, pp. 80–91, 2010.

[49] A. Greco, G. Valenza, M. Nardelli, M. Bianchi, L. Citi, and E. P. Scilingo, "Force–velocity assessment of caress-like stimuli through the electrodermal activity processing: Advantages of a convex optimization approach," *IEEE Transactions on Human-Machine Systems*, vol. PP, no. 99, pp. 1–10, 2016.

[50] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 2, pp. 153–158, 1997.

[51] J.-C. Wang, J.-F. Wang, K. W. He, and C.-S. Hsu, "Environmental sound classification using hybrid svm/knn classifier and mpeg-7 audio low-level descriptor," in *Neural Networks, 2006. IJCNN06. International Joint Conference on*. IEEE, 2006, pp. 1731–1735.

[52] M. Kearns and D. Ron, "Algorithmic stability and sanity-check bounds for leave-one-out cross-validation," *Neural Computation*, vol. 11, no. 6, pp. 1427–1453, 1999.

[53] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, no. June, pp. 271–274, 1998.

[54] A. Lanata, G. Valenza, and E. Scilingo, "A novel EDA glove based on textile-integrated electrodes for affective computing," *Medical and Biological Engineering and Computing*, vol. 50, pp. 1163–1172, 2012.

[55] A. Greco, A. Lanata, L. Citi, N. Vanello, V. Gaetano, and E. P. Scilingo, "Skin admittance measurement for emotion recognition: A study over frequency sweep," *Electronics*, 2016.

[56] P. Lang, M. Greenwald, M. Bradley, and A. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.