Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity

Article *in* IEEE Transactions on Affective Computing · February 2019

DOI: 10.1109/TAFFC.2019.2901673

5 authors, including:

Jainendra Shukla
Indraprastha Institute of Information Technology
24 PUBLICATIONS **102** CITATIONS
SEE PROFILE

Miguel Barreda-Ángeles
Vrije Universiteit Amsterdam
33 PUBLICATIONS **144** CITATIONS
SEE PROFILE

Joan Oliver
Instituto de Robótica para la Dependencia, Sitges, Spain
9 PUBLICATIONS **58** CITATIONS
SEE PROFILE

G. C. Nandi
Indian Institute of Information Technology Allahabad
151 PUBLICATIONS **1,439** CITATIONS
SEE PROFILE

Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity

Jainendra Shukla, Miguel Barreda-Ángeles, Joan Oliver, G. C. Nandi, *Senior Member IEEE* and Domènec Puig

Abstract—Electrodermal activity (EDA) is indicative of psychological processes related to human cognition and emotions. Previous research has studied many methods for extracting EDA features; however, their appropriateness for emotion recognition has been tested using a small number of distinct feature sets and on different, usually small, data sets. In the current research, we reviewed 25 studies and implemented 40 different EDA features across time, frequency and time-frequency domains on the publicly available AMIGOS dataset. We performed a systematic comparison of these EDA features using three feature selection methods, Joint Mutual Information (JMI), Conditional Mutual Information Maximization (CMIM) and Double Input Symmetrical Relevance (DISR) and machine learning techniques. We found that approximately the same numbers of features are required to obtain the optimal accuracy for the arousal recognition and the valence recognition. Also, the subject-dependent classification results were significantly higher than the subject-independent classification for both arousal and valence recognition. Statistical features related to the Mel-Frequency Cepstral Coefficients (MFCC) were explored for the first time for the emotion recognition from EDA signals and they outperformed all other feature groups, including the most commonly used Skin Conductance Response (SCR) related features.

Index Terms—EDA, Emotion Recognition, Arousal, Valence, Feature Extraction, Feature Selection.

1 Introduction

Robust information about the emotional state of a user is key to provide an empathetic experience during Human-Machine Interaction [1] or Human-Robot Interaction [2] [3]. In this regard, one important psychophysiological source of information is user's Electrodermal Activity (EDA), which refers to the conductivity of the skin, usually measured in palmar sites, and which can reflect changes in cognitive and emotional processes, such as cognitive effort or emotional arousal [4]. EDA signal, a non-stationary signal, is an aggregate of two different components: a tonic component, which accounts for the general levels of the conductivity of the skin, and whose levels vary slowly over time, and a phasic component, which manifests as sharper peaks over the tidal drift of the tonic EDA, which in general result from momentary sympathetic activation when arousing stimuli are present, although they also occur spontaneously in some individuals [5] [4] [6]. The tonic component of EDA is commonly called skin conductance level (SCL), while the short-term phasic responses are called skin conductance responses (SCRs). Thus, the EDA signal typically shows relatively rapid increments in its levels, followed by much slower decrements back to the baseline level (see Fig. 1). EDA mean levels usually range between 2 and 20μS, and vary within a range between 1 and 3μS for different individuals; the typical rise time from a valley to a peak is about 1 to 3 seconds and the typical recovery time of half of the SCR amplitude is between 2 and 10 seconds [4].

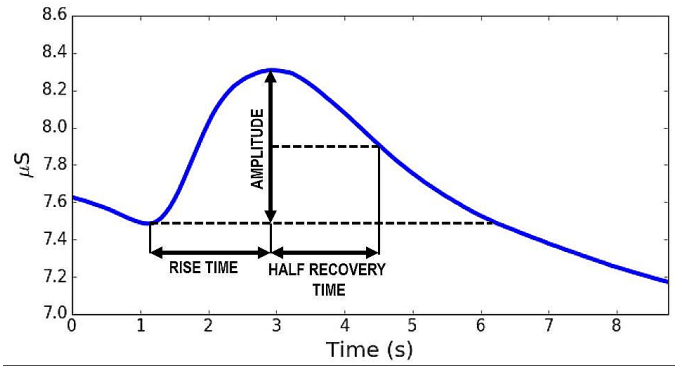

Fig. 1. Example EDA Signal: Adapted from [4]

Due to its low cost and easy-to collect nature, EDA measurement has been commonly applied in the context of psychological research. Recently, researchers have begun to explore the adequacy of machine learning algorithms for predicting user's mental states from EDA when interacting with diverse technologies. In this sense, EDA measurement may be especially useful when users have some kind of limitation at providing self-reported information, as is the case in a wide range of health conditions (e.g. [2] [3]). However, to date, there has not been a systematic exploration of the predictive power of using different combinations of EDA features. The goal of the current research is to contribute in filling this gap by examining the performance of a broad set of EDA features for emotion recognition.

The remainder of this article is structured as follows:

J. Shukla is with Department of Computer Science and Engineering, Indraprastha Institute of Information Technology Delhi (IIIT-D), New Delhi 110020, India. E-mail: jainendra@iiitd.ac.in
J. Shukla and J. Oliver are with the Instituto de Robótica para la Dependencia, Sitges, Barcelona, Spain E-mail: jshukla, joliver@institutorobotica.org
M.Barreda-Ángeles is with Eurecat, Centre Tecnològic de Catalunya, Barcelona, Spain
G. C. Nandi is with Department of Robotics & Artificial Intelligence, Indian Institute of Information Technology Allahabad (IIIT-A), Allahabad, India
J. Shukla, J. Oliver and D. Puig are with Intelligent Robotics and Computer Vision Group, Universitat Rovira i Virgili, Tarragona, Spain

Section 2 reviews feature extraction and feature selection methods used for emotion classification from EDA. In Section 3, we describe the methods used for extraction, selection, and classification of the EDA features along with the employed public dataset, and the processing of EDA signal for features comparison. We provide the results in Section 4, and they are discussed in Section 5. Finally, Section 6 concludes the article by summarizing this work while highlighting its limitation and proposing future steps.

2 Related Work

Emotion recognition from EDA has been commonly used for the assessment of user's experience in a variety of contexts such as recreational and serious games [7], driving [8], or patient-robot interaction [9]. Previous research has explored the predictive power of a diverse set of EDA features of different types, including time domain, frequency domain, and time-frequency domain features.

2.1 EDA Features

2.1.1 Time Domain Features

Regarding time domain features, some of the most usually considered features are statistical parameters of the signal during a relatively long recordings, including features such as signal mean value, standard deviation, kurtosis, or skewness (e.g. [10] [11] [12]). In other cases, researchers have focused on event-related features of EDA. Event related features refer to the attributes of the short-term responses, a few seconds after the presentation of a certain stimulus (e.g. images or sounds), such as the presence or absence of a SCR. In this sense, SCRs can be automatically detected in longer time windows and features can be extracted from them. This involves the definition of thresholds, in order to restrict the analysis only to non-negligible responses, and discarding those small changes in the signal that does not reach the thresholds and therefore cannot be considered as SCRs. In this sense, a traditional threshold for SCR amplitude has been $0.05\mu S$ [5]. Examples of event-related aspects of EDA considered in previous studies are SCR amplitude, SCR peaks count, mean SCR rise time, or the sum of SCR areas [13] [8] [14] [15] [11].

Other types of time domain features that have been successfully applied for emotion recognition in other types of physiological signals, such as Higher Order Crossing (HOC) features [16] [17], or Hjorth features from EEG signals [18] [19], might also have predictive potential in the case of EDA. However, to our knowledge, these features have not been previously applied to EDA signals.

2.1.2 Frequency Domain Features

Compared to time-domain features, fewer research has focused on the predictive power of EDA features related to the frequency domain, although the transient characteristics of the EDA signal can be better understood by analyzing the frequency domain representation of the signal. Indeed, the frequency domain analysis has shown superior capability for the gradient component's detection of individual SCR over traditional amplitude analysis [20]. Fast Fourier Transform (FFT), Short-time Fourier Transform (STFT) and power spectral density (PSD) estimation using Welch's method have been the algorithms most commonly used to obtain the frequency domain representation of the signal. Due to the different rate of physiological processes, EDA signals vary significantly with the frequency [21]. Hence, frequency oscillations of EDA signals can be divided into different frequency subbands to analyze it in more detailed manner [10] [21]. Indeed, previous research has considered statistical aspects (variance, range, signal magnitude area, skewness, kurtosis, harmonics summation) and spectrum power of five frequency bands, as well as their minimum, maximum, and variance [15] [8] [22] [10] [21] [23].

2.1.3 Time-Frequency Domain Features: Wavelets

Since EDA exhibits non-stationary behavior, wavelets have been found suitable for modeling EDA activity [24].

Discrete Wavelet Transform

When the wavelets are discretely sampled, the wavelet transform is known as Discrete Wavelet Transform (DWT). Denoised DWT wavelet coefficient features have been used for emotional state classification in patient-robot interaction [9].

Stationary Wavelet Transform

On the other hand, Stationary Wavelet Transform (SWT) also present some advantages for the analysis of EDA. SWT is redundant, linear and hence shift invariant in comparison to the DWT [25]. SWT also provides better sampling rates in the low-frequency bands compared with a standard DWT [25]. The SWT technique has been successfully applied to denoising of EDA signals with high efficacy and less computational complexity [26].

2.1.4 Mel-Frequency Cepstrum Features

The EDA signal can be characterized by a sequence of overlapping rapidly varying phasic SCRs overlying a slowly varying tonic activity (i.e., SCL). This superposition complicates the proper decomposition of the Skin Conductance (SC) data and hence limits the ability of classical methods for the assessment of SCRs [5]. Sudomotor nerve activity can be considered as a driver of EDA and it consists of a sequence of mostly distinct impulses (i.e., sudomotor nerve bursts). These bursts trigger the specific impulse response (i.e., SCRs) and thereby the SC can be modeled by driver-impulse response (IR) convolution. As a result of this process, SC can be represented by [27]:

$$\begin{aligned} SC &= SC_{tonic} + SC_{phasic} \\ &= (Driver_{tonic} + Driver_{phasic}) * IR \end{aligned} \tag{1}$$

In this model, SC is considered the output of the skin system driven by an excitation sequence of sudomotor nerve bursts, but such convolution of the response and the drivers cannot be easily separated in the time domain. Cepstrum analysis (CA) is an important technique for analyzing similar models of the speech signal [28].

The cepstrum of a discrete-time signal is the inverse discrete-time fourier transform of the logarithm of the magnitude of the discrete-time Fourier transform (DTFT) of the signal and is given by:

$$c[n] = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \log |X(e^{i\omega})| e^{i\omega n} d\omega \tag{2}$$

where $X(e^{i\omega})$ is the DTFT of the signal.

CA has been used successfully to isolate the basic waveform and the excitation function of the physiological signals such as electrocardiogram [29], electroencephalogram [30] and even EDA [21]. In [21], while analyzing EDA signals using CA, it was concluded that CA could be useful for analysis of superimposed EDA signals given its ability to magnify small amplitude variations. It suggests possible advantages of using Mel-frequency cepstral coefficients (MFCC), a new type of cepstrum representation based on the weighted cepstrum distance measures [31], that is widely established for many pattern recognition problems related to speech signals [28], as a feature vector of EDA signals. However, to our knowledge, this has not been addressed in previous studies.

In summary, several studies have explored the predictive potential for emotion classification of a broad set of EDA features; however, no previous study has tackled a systematic comparison of them, and the reliability of the recognition results remains challenging. Moreover, processing high dimensional data demands significant computational and space complexity. Therefore, extracting emotion information from high dimensional EDA data can be challenging, especially if the processing is to be done online. Furthermore, many of the EDA features can be irrelevant for the emotion classification or they might be redundant. Hence, it is important to automatically identify meaningful smaller subsets of these EDA features to achieve efficient emotion recognition from EDA signals, which stresses the need for using effective feature selection methods.

2.2 Feature Selection Methods

Most of the studies on EDA features have either not employed any algorithm for feature selection (FS) or have just applied data reduction techniques such as principal component analysis (PCA) prior to the classification [21] [9]. PCA is not a beneficial transformation for pattern recognition problems as the objective of PCA is not related to the automated target recognition [32]. Also, the recognition reported with PCA cannot be easily generalized to a general set-up [32]. FS methods can be generally divided into classifier-dependent ('wrapper' and 'embedded' methods), and classifier-independent ('filter' methods) categories [33]. Wrapper and embedded methods are computationally expensive and both use quite a strict model structure assumptions and hence may produce classifier specific feature subsets. In contrast, filter methods produce generic feature subsets, as they are model independent. Filter methods also take into account how potentially useful a feature or feature subset may be when used in a classifier [33]. In [33], the authors provide a comprehensive review of information based filter FS algorithms. Furthermore, they propose three desirable characteristics of an information-based selection criterion, which are [33]:

1) Whether it includes reference to a conditional redundancy term?
2) Whether it balances the relevance and the redundancy terms?
3) Whether it uses a low dimensional approximation?

The research conducted in [33] found that only three FS algorithms, Joint Mutual Information (JMI), Conditional Mutual Information Maximization (CMIM), and Double Input Symmetrical Relevance (DISR), show these desirable characteristics of an information-based selection criterion, and, therefore, we focus on these three methods.

2.2.1 Joint Mutual Information

Joint Mutual Information (JMI) criterion [34] provides the best trade-off in terms of accuracy, stability, and flexibility [33]. JMI focuses on increasing complementary information between features. The JMI score for feature X_k is

$$J_{JMI}(X_k) = \sum_{X_j \in S} I(X_k X_j; Y) \tag{3}$$

This is the information between the target Y and a joint random variable $X_k X_j$, defined by pairing the candidate X_k with each feature X_j previously selected. The candidate feature X_k that maximizes this mutual information is chosen and added to the feature subset S. JMI offers two significant advantages [34]:

1) JMI can distinguish among features even when all of them have same mutual information (MI) and
2) JMI can eliminate the redundancy in the features when one feature is a function of other features.

2.2.2 Conditional Mutual Information Maximization

Conditional Mutual Information Maximization (CMIM) [35] is a versatile filter measure and is a good measure for general FS problems [36]. The CMIM criterion for each feature X_k is measured as

$$J_{CMIM}(X_k) = \min_{X_j \in S} [I(X_k : Y | X_j)] \tag{4}$$

$I(X_k : Y|X_j)$ is the conditional mutual information between candidate X_k and the target Y given X_j. The candidate feature X_k that minimizes this conditional mutual information is chosen and added to the feature subset S which means that it carries information about the class that is not already captured by the features in the selected set. CMIM can properly identify truly redundant features and noisy features, and gives preference to informative, uncorrelated features [36].

2.2.3 Double Input Symmetrical Relevance

Double Input Symmetrical Relevance (DISR) [37] is a normalized variant of JMI. DISR takes into consideration the variable complementary and a lower bound on the mutual information. It uses the following modification of the JMI criterion:

$$J_{DISR}(X_k) = \sum_{X_j \in S} \frac{I(X_k X_j; Y)}{H(X_k X_j Y)} \tag{5}$$

DISR criterion motivates the selection of a complementary variable of an already selected one with a higher probability.

2.3 Problem Statement

It is not known which features are most appropriate for emotion recognition from EDA and previous works have made limited contributions on a systematic comparison of EDA features. Hence, the research goals of our work are as follows:

1) To provide an inclusive review of EDA features for emotion recognition.
2) To provide a first-ever systematic comparison of features on one database using multiple FS methods.
3) To identify the most significant EDA features for emotion recognition.

3 METHODS

In order to tackle the above research goals, we described a wide set of features that can be extracted from EDA signals in different domains. We then extracted these features from EDA signals of a publicly available, annotated dataset, the AMIGOS dataset [38], and implemented a systematic FS and comparison process to determine the most significant features. The whole process is described below.

3.1 Features Set

A literature review was conducted in order to determine a complete set of features available for emotion classification from EDA, as well as other types of features that have proven effective for this purpose when applied to other types of psychophysiological signals, even if they have not been applied to EDA. We retrieved the relevant publications by performing an online search of the PubMed, IEEE Xplore, ScienceDirect research databases for studies published in English until date. The keywords used for the search were combinations of relevant words such as EDA, Electro-dermal Activity, Emotion etc. Based on the their titles and the abstracts, we manually identified the studies discussing EDA features and selected only original studies published in journals or conferences. We created a data extraction spreadsheet for the collection of different EDA features from these studies. Furthermore, we discarded the duplicate references by manually reviewing the feature list and the references. Finally, 25 papers were used for exploring and implementing the 40 different EDA related features. The complete list of selected features is presented in Table 1.

3.2 AMIGOS Database

The AMIGOS dataset [38] is a publicly available dataset containing, among other multimodal data, measures of EDA from two experiments: one experiment in which participants watched short ($<250s$) emotional videos (40 participants), and another experiment in which participants (alone or in groups of four) watched longer ($>14min$) videos able to elicit diverse emotional states (37 participants, 17 of them in individual setting and 20 in groups). Among others, the dataset includes annotations for emotional arousal and valence of participants in both experiments, provided by three external observers that visually inspected frontal videos of participants' faces during the viewing, and provided an annotation for every 20-second segment of the viewing. A total of 12,580 video segments were annotated this way (340 segments by 37 participants, from both the experiments with short and long videos). The arousal and valence scales used for these annotations were continuous and ranged from -1 (low arousal or valence) to +1 (high arousal or valence), and the agreement between annotators was very good for both the variables (Cronbach's α = 0.96 for arousal, and Cronbach's α = 0.98 for valence).

3.3 Feature Extraction

Table 2 describes the dimensionality of each feature type for each participant. A feature matrix was generated from the EDA data of 340 annotated segments for each of the two annotated variables (arousal and valence) for each participant. Features were extracted from all the three domains leading to a total of 621 features. We z-normalized the features to have mean 0 and to have a standard deviation of 1. The problem of singularities may occur for FS methods [17]. Hence, we removed all almost identical features which produced a correlation coefficient higher than .98. In what follows, we explain the process that we used to extract each type of features.

3.3.1 Extraction of Time Domain Features

Event Related Features

We used the process described in [11] [8] [13] to extract the following event-related features: SCR Amplitude, SCR peak count, mean SCR amplitude, mean SCR rise time, sum of SCR peaks amplitudes, sum of SCR rise times, area under the curve of SCRs and sum of SCR areas.

Statistical Features

Based on the methods described in [10] [44], we extracted the following statistical features: power, mean, standard deviation, kurtosis, skewness, mean of the 1st difference, and mean of the 2nd difference.

Hjorth Features

In addition to the commonly used event-related and statistical features from EDA, we also extracted three Hjorth features, following the methods described in [47]. In this regard, for a given time series signal $X(n)$ of the EDA signal, we extracted:

- Activity (A):

$$A_X = \sum_{n=1}^{N} (X(n) - \mu)^2 \tag{6}$$

- Mobility (M):

$$M_X = \sqrt{var(\dot{X(n)})var(X(n))} \tag{7}$$

- Complexity (C):

$$C_X = M(\dot{X(n)})M(X(n)) \tag{8}$$

where μ is the mean, $var(X(n))$ is the variance, $var(\dot{X(n)})$ is the derivative of the variance of the EDA signal $X(n)$.

HOC

We also extracted the HOC features of given EDA time-series signals using the method described in [48]. When a specific sequence of filters is applied iteratively to a time

TABLE 1
EDA features used in the research

	Features	Parameters	Description	References
Time Domain	SCR Features	peakCount	SCR amplitude peak counts	[8], [11], [39], [40]
		meanpeakAmplitude	SCR amplitude mean	[11], [13], [14], [15], [22], [39], [40], [41]
		Mean Rise Time	SCR amplitude mean rise time	[8], [13], [15], [39]
		Sum Peak Amplitude	SCR amplitude summation	[8], [13], [41]
		Sum Rise Time	SCR amplitude rise time summation	[8], [13], [15], [22]
		Sum Areas	Sum of estimated areas of orienting responses	[8], [40], [42]
		auc	Area under curve	[40]
	Statistical Features	meanEDA	Mean of signal	[8], [10], [12], [13], [21], [39], [40], [43], [44], [45]
		stdEDA	Standard deviation of signal	[8], [13], [21], [39], [40], [43], [44], [45]
		kurtEDA	Kurtosis of signal	[21], [44], [45], [46]
		skewEDA	Skewness of signal	[21], [44], [46]
		meanDerivative	Mean of 1st order derivatives	[10], [11], [39], [44]
		meanNegativeDerivative	Mean of negative values of 1st order derivatives	[22], [44]
	Hjorth Features*	Activity	Variance of the signal	[18], [19], [47]
		Mobility	Square root of variance of the first derivative of the signal divided by variance of the signal	[18], [19], [47]
		Complexity	Mobility of Mobility	[18], [19], [47]
	Higher Order Crossings*	HOC	Sequence of zero-crossings of a specific sequence of filtered signal	[16], [17], [48]
Frequency Domain	Statistical Features	SMA	Signal Magnitude Area	
		meanEDA	Mean of signal	[21]
		stdEDA	Standard deviation of signal	[21]
		signalRange	Range of Signal	[21]
		harmonicsSummation	Summation of FFT Harmonics	[21]
		kurtEDA	Kurtosis of signal	[21]
		skewEDA	Skewness of signal	[21]
		signalEnergy	Energy of the signal	[8], [15]
	Band Power	SpectralPower [0.05-0.5 Hz], $\delta = 0.1$	5 spectral power in the [0-0.5]Hz bands	[10], [22]
		minSpectralPower	Minimum of spectral band powers	[21], [23]
		maxSpectralPower	Maximum of spectral band powers	[21], [23]
		varSpectralPower	Variance of spectral band powers	[21], [23]
Time-Frequency Domain	Wavelet Features	Energy distribution*	Energy percentage for wavelet levels	
		energyWavelet*	Energy for wavelet levels [Absolute]	
		entropyWavelet*	Entropy for wavelet levels	
		rmsWavelet*	Root mean square for wavelet coefficients	
		coefficients	Approximation and detail coefficients	[9]
	MFCC*	mfccCoefficients	MFCC coeffcients for all frames	[21], [31]
		meanMFCC	Mean of MFCC coefficients	[21], [31]
		stdMFCC	Standard deviation of MFCC coefficients	[21], [31]
		medianMFCC	Median of MFCC coefficients	[21], [31]
		kurtMFCC	Kurtosis of MFCC coefficients	[21], [31]
		skewMFCC	Skewness of MFCC coefficients	[21], [31]

TABLE 2
Feature Vector Dimension

Domain	Feature Vector	Number of Features
Time Domain	Skin conductance response (SCR) Related	7
	Statistical Features	8
	Hjorth Features	2
	Higher Order Crossings	5
Frequency Domain	Statistical Features	8
	Band Power	9
Time-Frequency Domain	Discrete Wavelet Transform (DWT) Coefficients	56
	Stationary Wavelet Transform (SWT) Features	40
	Mel-frequency cepstral coefficients (MFCC)	481
	Statistical features related to MFCC	5
Total		**621**

series Z_t, the corresponding sequence of zero-crossings is known as HOC sequence and is represented by:

$$\Im_k\{Z_t\} = \nabla^{k-1} Z_t \tag{9}$$

∇ is iteratively applied backward difference operator $\nabla Z_t \equiv Z_t - Z_{t-1}$ and the order k is selected in range of $1-50$ that would result in the maximum classification rate of the given signals [16]. The desired simple HOC are then obtained by counting the symbol changes D_k in $\Im_k\{Z_t\}$. HOCs are then used to construct the feature vector $\mathrm{FV}_{\mathrm{HOC}}$ as follows:

$$\mathrm{FV}_{\mathrm{HOC}} = [D_1, D_2, \ldots, D_L], \quad 1 < L \leq J \tag{10}$$

In order to determine the most suitable order for the HOC features, we performed an iterative classification step. We computed the classification rate for several orders of HOC features using stratified 10-fold cross-validation. The classification of the HOC features data was performed by means of quadratic discriminant analysis with diagonal covariance estimates (i.e. Naive Bayes). Figure 2 shows the plot of the HOC order vs corresponding classification rate for arousal recognition and for valence recognition. As it is clear from the plot, HOC gets its highest classification rate at order value 5 for both the arousal and the valence recognition, hence we chose the HOC order as 5 for the AMIGOS dataset.

3.3.2 Extraction of Frequency Domain Features

In order to extract the frequency domain features, the recommended frequency range of EDA signal ($0.05-0.50Hz$) was split into five bands following suggestions from [10] [21]. The features extracted from the resulting frequency domain representation of the EDA signal are a set of statistical features (variance, range, signal magnitude area, skewness, kurtosis, harmonics summation) and the spectrum power of five frequency bands, their minimum, maximum, and variance [15] [8] [22] [10] [21] [23].

3.3.3 Extraction of Time-Frequency Domain Features

Discrete Wavelet Transform

Based on a previous work [9], we used the detail coefficients (equation 13) of the DWT signal as EDA features. Wavelet analysis of a signal consists of translations ($k \in Z$) of the *father wavelet* $\phi(t)$ *(scaling function)* and dilations and translations ($k \in Z$, $j \in Z$) of the *mother wavelet* $\psi(t)$. The wavelet series representation of a signal $x(t)$ is then:

$$x(t) = \sum_j c_{j,k}\phi_k(t) + \sum_j \sum_k d_{j,k}\psi_{j,k}(t) \tag{11}$$

where $(c_{j,k})$ are the *approximation coefficients* and $(d_{j,k})$ are the *detail coefficients* of the wavelet coefficient set and are calculated as follows:

$$c_{j,k} = \langle x(t), \phi_{j,k}(t) \rangle \tag{12}$$

$$d_{j,k} = \langle x(t), \psi_{j,k}(t) \rangle \tag{13}$$

In DWT, wavelet acts as a band-pass filter from signal processing point of view, where the scaling and wavelet functions serve as low pass (h[n], Equation 14) and high pass filters (g[n], Equation 15), respectively.

$$\phi(t) = \sum_n h[n]\sqrt{2}\phi(2t-n) \tag{14}$$

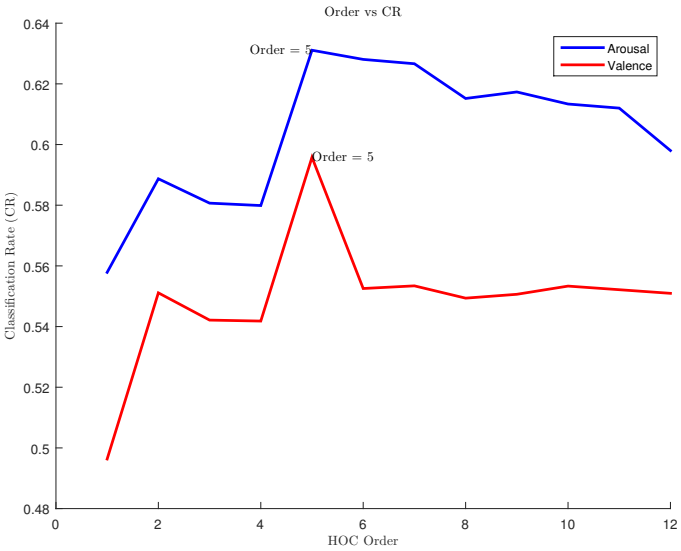

Fig. 2. HOC order vs classification rate

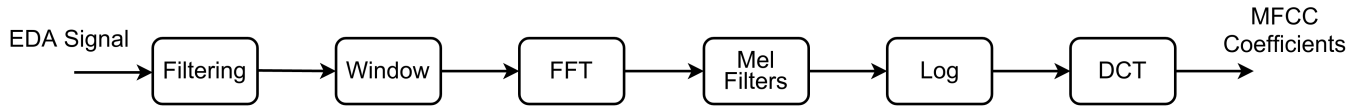

Fig. 3. MFCC Feature Extraction

$$\psi(t) = \sum_n g[n]\sqrt{2}\psi(2t - n) \tag{15}$$

where,

$$g[n] = h[2N - 1 - n] \tag{16}$$

When applied, the above decomposition halves the temporal resolution and doubles the frequency resolution. The above procedure can be applied iteratively for multilevel decomposition of the signal. Wavelet decomposition levels correspond to different frequency bands and this correspondence is based on the sampling frequency of the signal.

Stationary Wavelet Transform

To obtain the SWT of the given EDA signal, we used the process described in [25]. The basic DWT algorithm described above can be modified to obtain SWT of given EDA signal. We applied the low-pass and the high-pass filters as described for DWT, but without any decimation. We extracted the following SWT based features:

- Wavelet Energy: $E(j) = \sum_{i=1}^{j} D_i(n)^2$
- Wavelet Decomposition Energy: The percentage of energy corresponding to the approximation and to the details coefficients.
- Wavelet Entropy: $E(j) = -\sum_{i=1}^{j} D_i(n)^2 log(D_i(n)^2)$
- Wavelet Root Mean Square (RMS): $RMS(j) = \sqrt{\frac{1}{N}\sum_{i=1}^{j} |D_i(n)|^2}$

where $D_i(n)$ are the detail coefficients of the EDA signal.

3.3.4 Extraction of MFCC Features

The process applied for extracting MFCC features is illustrated in figure 3 and is explained below:

1) The EDA signal was filtered to remove the motion artifacts using the sophisticated SWT based filtering method [26].
2) A Hamming window was applied on the filtered EDA signal to enable the analysis over short window durations. Given a sampling frequency of f, the recommended value of frame size (N) is, $N = 2 \times f$ and of overlapping window duration (M) is, $M = 0.5 \times f$. The sampling frequency for EDA signals from the AMIGOS dataset is 128 Hz. While the frequency of the EDA signals is very high, the latency of the gradual changes in principle EDA components against an elicited stimulus is between 1.0 and 3.0s [4]. Keeping this in mind, we took the overlapping window duration M as 0.5 seconds, so as to cover the starting segment of responses for specific stimuli occurring at 1.0 seconds. Moreover, since we are analyzing a 20 seconds EDA segment of the dataset, we could not select a value of $2 \times f = 256$ seconds for the value of N. Hence, we decided to make 10 equal windows for the EDA signal and thereby chose the value of N as 2 seconds.
3) For each window, the frequency spectrum was obtained by applying a FFT.
4) The frequency spectrum was then mapped onto the Mel-scale through Mel-filters to obtain the Mel-spectrum. The Mel-scale mapping from the actual frequency f can be given as follows:

$$f_{mel} = 2595 \log_{10}(1 + \frac{f}{700}) \tag{17}$$

5) Next we took the log of the Mel-spectrum values.
6) Finally, CA was required as per equation 2 on the Mel-spectrum to obtain the MFCC features. While applying equation 2, the logarithm computation was committed, since it was computed in the previous processing step 5. Also, instead DTFT, the discrete cosine transform was applied because the absolute value of the Mel-spectrum is real and symmetric. Hence, for a windowed frame of EDA signal the final MFCC coefficients $C[n]$ are computed by

$$C[n] = \frac{1}{R}\sum_{r=1}^{R} \log(MF_m[r]).\cos\left[\frac{2\pi}{R}\left(\frac{r+1}{2}\right)n\right] \tag{18}$$

where R is the number of Mel-filters, $MF_m[r]$ is the Mel-spectrum of the frame r.

For the purposes of MFCC extraction, we selected only the last 13 components, since the rest of them carry little information. For a given EDA signal, the number of MFCC coefficients obtained via the above process resulted in $13 \times Num_{frames}$, where $Num_{frames} = ceil(((length(EDA_{signal}) - N)/M))$.

3.4 Feature Selection

As explained in section 2.2, among the several information theoretic FS methods available only JMI, CMIM and DISR satisfy the three desirable characteristics of an information based selection criterion [33]. Hence, we applied JMI, CMIM and DISR for selecting meaningful features from EDA. We used more than one FS method because the results obtained from multiple methods are more robust [49]. We applied each of these three FS methods for each participant individually and collectively for ALL participants after performing the scaling and discretization on the extracted EDA features.

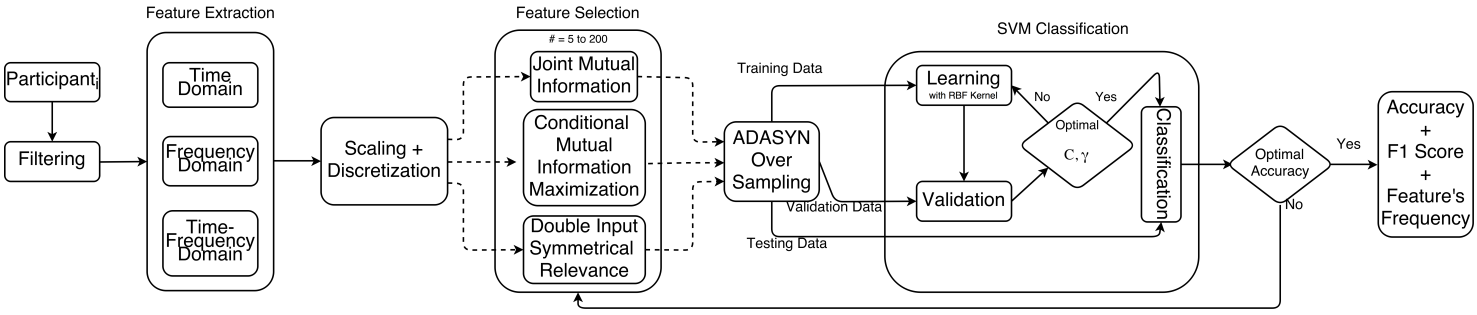

Fig. 4. Block diagram of supervised classification system

The formulas for applying the JMI, CMIM and DISR methods are given by the equations 3, 4 and 5 respectively. We evaluated the classifier performance by selecting the top n features as given by the FS algorithm. We varied the value of n between 5 to 200. We chose the upper value of 200 to check approximately one-third values of the feature vector (total number of features are 621 as given in table 2).

3.5 Classification

Since we intended to analyze temporal evolution of arousal and valence from EDA signals, we employed the external affect annotations of the AMIGOS dataset. We aggregated the ratings from all the three annotators to form a single rating value of more significant meaning for each annotated video segment. Since the valence and arousal scales were continuous and ranged from -1 (low arousal) to 1 (high arousal), we categorized rating values to LOW or HIGH valence/arousal depending on whether the ratings were less or greater than the mean value of the scale 0.00.

Based upon the above categorization, 9886 samples were assigned to class LOW and 2694 to class HIGH for arousal labeling, and 9566 were assigned to class LOW and 3014 to class HIGH for valence labeling. This provided an imbalanced two-class dataset. Hence, we used adaptive synthetic (ADASYN) sampling method to improve class balance towards equally-sized classes. ADASYN uses a weighted distribution for different minority class samples according to their level of difficulty in learning and it generates more synthetic data for minority class samples that are harder to learn in comparison to those minority samples that are easier to learn [50]. Adaptively generating synthetic data samples in this manner reduces the bias introduced by the imbalanced data distribution.

The overall methodology of our recognition system is illustrated in figure 4. The sample set was partitioned for each of the 37 participants individually (containing 340 samples for each participant) and collectively for ALL participants using the whole dataset (12,580 samples). The ADASYN method was applied for each data partition to remove the class imbalance and the data was then divided into the ratio of $70:15:15$ for training, validation and testing respectively. We evaluated each of the listed FS methods for each subject individually based on the classification accuracy with a support vector machine (SVM) classifier and radial basis function (RBF) kernel. We employed grid search and 3-fold cross validation method to determine the optimal regularization parameter C and also to determine the free parameter γ of the Gaussian RBF. We applied SVM since it has been reported to provide the best classification accuracy during the recognition of affective states from physiological cues [51] [52] [53].

4 Results

4.1 Optimal Number of Features

Table 3 presents the optimal accuracies for arousal recognition, average F1 score (F1, average of score for both classes) and the optimal number of features in tabular form across all the 37 subjects and the three FS methods. The optimal accuracies are the highest accuracies obtained at the optimal number of features. The table also presents the average results for all 37 subjects across all three FS methods, as well as the subject-independent (ALL) classification results. Table 4 shows the same information for valence recognition.

We tested if the values of F1-scores provided by the different selection algorithms are significantly higher than 0.5 (at the $p<.05$ level). In the case of arousal detection, the JMI ($M = 0.58; SD = 0.17; t(36) = 2.78; p = .009$), the CMIM ($M = 0.63; SD = 0.1; t(36) = 7.92; p<.001$), and the DISR ($M = 0.64; SD = 0.11; t(36) = 7.49; p = <.001$) algorithms provided mean values of F1-scores that were significantly higher than 0.5. Also in the case of valence, the JMI ($M = 0.61; SD = 0.1; t(36) = 6.49; p<.001$), CMIM ($M = 0.59; SD = 0.13; t(36) = 4.32; p<.001$), and DISR ($M = 0.63; SD = 0.09; t(36) = 8.85; p<.001$) algorithms provided values significantly higher than 0.5 for F1-scores.

In order to provide insight on how the performance of the subject-dependent classification differs from the subject-independent classification, we also compared the results of subject-independent (ALL) classification with subject-dependent (37 individual subjects) across all three FS methods for arousal and valence recognition. In the case of arousal, compared to the subject-independent (ALL) recognition, subject dependent recognition provided significantly higher accuracy (JMI: $t(36) = 17.1888, p<0.001$, CMIM: $t(36) = 18.4018, p<0.001$, DISR: $t(36) = 20.5805, p<0.001$), while having a significantly lower optimal number of features (JMI: $t(36) = -8.184, p < .001$, CMIM: $t(36) = -5.0724, p < .001$, DISR: $t(36) = -11.0016, p < .001$).

A similar pattern was observed for valence recognition, including a higher accuracy (JMI: $t(36) = 17.389, p<0.001$, CMIM: $t(36) = 18.3873, p<0.001$, DISR: $t(36) =$

TABLE 3
Accuracy, F1 Score, Optimal Number of Features for Different FS Methods Listed for Each Subject for Arousal Recognition

Participant	JMI			CMIM			DISR			Average		
	A*	F1#	F+	A*	F1#	F+	A*	F1#	F+	A*	F1#	F+
1	77.5	0.66	37	76.5	0.68	145	76.5	0.66	63	76.8	0.7	81.7
2	80.4	0.49	144	81.4	0.5	165	83.3	0.61	127	81.7	0.53	145.3
3	88.2	0.54	13	92.2	0.58	26	88.2	0.54	153	89.5	0.53	64
4	82.4	0.6	77	80.4	0.63	81	80.4	0.53	70	81.1	0.57	76
5	85.3	0.74	159	86.3	0.71	171	86.3	0.79	60	86	0.73	130
6	88.2	0.72	34	88.2	0.75	136	86.3	0.53	48	87.6	0.67	72.7
7	73.5	0.63	92	79.4	0.62	60	77.5	0.64	6	76.8	0.6	52.7
8	86.3	0.57	137	89.2	0.55	67	88.2	0.54	105	87.9	0.57	103
9	93.1	0.6	36	93.1	0.59	27	93.1	0.59	46	93.1	0.6	36.3
10	90.2	0.56	131	91.2	0.57	16	92.2	0.58	13	91.2	0.6	53.3
11	82.4	0.54	88	81.4	0.62	191	84.3	0.65	107	82.7	0.6	128.7
12	80.4	0.49	183	81.4	0.53	185	82.4	0.58	25	81.4	0.53	131
13	75.5	0.55	53	73.5	0.59	180	74.5	0.54	167	74.5	0.57	133.3
14	93.1	0.59	130	94.1	0.73	106	91.2	0.57	61	92.8	0.63	99
15	83.3	0.58	142	87.3	0.62	122	84.3	0.51	155	85	0.57	139.7
16	87.3	0.62	197	87.3	0.53	167	87.3	0.62	39	87.3	0.57	134.3
17	83.3	0.64	62	83.3	0.66	162	85.3	0.7	14	84	0.67	79.3
18	78.4	0.52	200	78.4	0.7	194	80.4	0.7	13	79.1	0.63	135.7
19	97.1	0.69	86	98	0.75	6	96.1	0.66	123	97.1	0.73	71.7
20	90.2	0.66	92	91.2	0.63	174	93.1	0.84	105	91.5	0.7	123.7
21	82.4	0.74	70	85.3	0.74	83	85.3	0.76	42	84.3	0.73	65
22	100	1	186	100	1	166	100	1	195	100	1	182.3
23	85.3	0.52	131	82.4	0.5	14	82.4	0.6	161	83.4	0.53	102
24	93.1	0.66	145	93.1	0.59	65	94.1	0.68	15	93.4	0.67	75
25	87.3	0.53	6	88.2	0.54	24	89.2	0.55	138	88.2	0.53	56
26	88.2	0.54	58	93.1	0.59	92	88.2	0.54	28	89.8	0.53	59.3
27	84.3	0.51	197	88.2	0.63	69	85.3	0.56	87	85.9	0.57	117.7
28	86.3	0.57	158	84.3	0.59	190	88.2	0.67	39	86.3	0.63	129
29	78.4	0.71	183	80.4	0.71	187	81.4	0.74	139	80.1	0.7	169.7
30	77.5	0.56	75	77.5	0.64	31	76.5	0.65	155	77.2	0.63	87
31	72.6	0.59	7	76.5	0.63	7	78.4	0.72	10	75.8	0.63	8
32	89.2	0.6	53	89.2	0.55	197	91.2	0.63	20	89.9	0.6	90
33	70.6	0.69	26	68.6	0.67	53	68.6	0.69	38	69.3	0.7	39
34	83.3	0.51	185	84.3	0.51	48	82.4	0.5	101	83.3	0.5	111.3
35	92.2	0.58	162	91.2	0.57	87	96.1	0.66	24	93.2	0.63	91
36	89.2	0.68	36	89.2	0.55	77	87.3	0.53	74	88.6	0.6	62.3
37	97.5	0.7	81	96.1	0.66	71	98	0.75	90	97.2	0.73	80.7
Average	85.2	0.61	104	86	0.63	104	86	0.64	77	**85.75**	**0.63**	**95.04**
ALL	65.1	0.65	187	64.6	0.65	158	62.6	0.62	176	**64.1**	**0.64**	**173.6**

A*– Classification Accuracy F1#– F1 Score F+– Optimal Number of Features

$21.8212, p<0.001$), and lower optimal number of features (JMI: $t(36) = -10.3046, p < .001$, CMIM: $t(36) = -4.7272, p < .001$, DISR: $t(36) = -6.31, p < .001$) for the subject-dependent recognition compared to the subject-independent (ALL) recognition.

A series of paired t-tests were conducted in order to determine if there was a significant difference between the average values (across the three FS algorithms) of accuracy for the prediction of arousal and valence, and the results showed that there was not, $t(36) = 1.79; p = .08$.

Finally, the outputs of the three FS algorithms in terms of accuracy and optimal number of features were compared using a series of repeated measures (within-subjects), analysis of variance (ANOVAs), in order to examine if there is a significant difference between them. Regarding arousal, no significant difference (at $p<.05$ level) was found between the three algorithms regarding their accuracy ($F(2, 105) = 0.1; p = .91$), or optimal number of features ($F(2, 105) = 1.07; p = .35$). Neither in the case of valence there were significant differences in terms of the accuracy ($F(2, 105) = 0.01; p = .99$), or optimal number of features ($F(2, 105) = 0.39; p = .68$). In other words, there is no evidence that one of the algorithms outperforms the others.

4.2 Significant Features

In order to identify the most frequently selected features by the different FS methods and the features with the most significant performance for arousal and valence recognition, we followed an approach similar to the one described in [17]. It involves the computation of the relative frequency of each of the feature types, which is obtained via the following process:

1) We first created a histogram of the feature occurrence, given the features selected for the optimal number of features across all the participants and the FS methods.
2) Then, in order to take into account the random assignment of the features, each bin of the histogram was normalized by dividing the occurrence of each feature type by the feature's cardinality (e.g. 5 in the case of HOC).

TABLE 4
Accuracy, F1 Score, Optimal Number of Features for Different FS Methods Listed for Each Subject for Valence Recognition

Participant	JMI			CMIM			DISR			Average		
	A*	F1#	F+	A*	F1#	F+	A*	F1#	F+	A*	F1#	F+
1	70.6	0.58	20	68.6	0.55	171	68.6	0.61	164	69.3	0.58	118.3
2	76.5	0.58	68	77.5	0.54	73	77.5	0.58	192	77.1	0.57	111
3	87.3	0.53	122	85.3	0.52	7	82.4	0.54	140	85	0.53	89.7
4	75.5	0.6	155	76.5	0.63	65	78.4	0.59	41	76.8	0.61	87
5	91.2	0.78	75	88.2	0.74	104	89.2	0.79	152	89.5	0.77	110.3
6	85.3	0.56	45	87.3	0.53	50	84.3	0.59	116	85.6	0.56	70.3
7	66.7	0.67	19	66.7	0.65	20	67.6	0.67	93	67	0.66	44
8	74.5	0.49	195	82.4	0.54	132	81.4	0.57	13	79.4	0.53	113.3
9	85.3	0.52	132	91.2	0.57	124	94.1	0.68	12	90.2	0.59	89.3
10	84.3	0.56	139	86.3	0.53	53	86.3	0.61	105	85.6	0.56	99
11	75.5	0.63	92	77.5	0.65	90	77.5	0.64	190	76.8	0.64	124
12	76.5	0.5	5	73.5	0.62	113	72.5	0.53	69	74.2	0.55	62.3
13	69.6	0.57	53	75.5	0.71	49	76.5	0.67	161	73.9	0.65	87.7
14	95.1	0.63	44	91.2	0.57	185	91.2	0.57	58	92.5	0.59	95.7
15	81.4	0.5	112	83.3	0.55	76	81.4	0.59	82	82	0.55	90
16	77.5	0.54	156	81.4	0.5	72	80.4	0.56	47	79.7	0.53	91.7
17	77.5	0.6	126	78.4	0.54	111	80.4	0.68	46	78.8	0.61	94.3
18	88.2	0.74	59	88.2	0.67	39	89.2	0.6	55	88.6	0.67	51
19	91.2	0.57	147	92.2	0.69	165	93.1	0.59	109	92.2	0.62	140.3
20	93.1	0.84	94	92.2	0.8	199	94.1	0.85	160	93.1	0.83	151
21	77.5	0.66	93	80.4	0.72	162	79.4	0.71	72	79.1	0.7	109
22	86.3	0.53	12	85.3	0.52	89	84.3	0.51	167	88.6	0.52	89.3
23	75.5	0.5	64	78.4	0.59	16	79.4	0.55	7	77.8	0.55	29
24	95.1	0.63	61	93.1	0.71	24	93.1	0.71	16	93.8	0.69	33.7
25	77.5	0.56	56	79.4	0.52	123	81.4	0.59	45	79.4	0.56	74.7
26	87.3	0.53	188	86.3	0.53	79	85.3	0.52	161	86.3	0.53	142.7
27	88.2	0.59	51	91.2	0.57	138	91.2	0.63	29	90.2	0.6	72.7
28	85.3	0.6	71	83.3	0.55	178	83.3	0.58	86	84	0.58	111.7
29	92.2	0.85	171	90.2	0.72	106	91.2	0.82	198	91.2	0.8	158.3
30	79.4	0.58	30	80.4	0.56	134	78.4	0.63	75	79.4	0.59	79.7
31	87.3	0.8	7	78.4	0.59	156	83.3	0.76	18	83	0.72	60.3
32	90.2	0.56	167	92.2	0.73	61	97.1	0.83	142	93.1	0.7	123.3
33	84.3	0.56	183	83.3	0.55	140	85.3	0.52	160	84.3	0.54	161
34	83.3	0.58	31	84.3	0.56	184	87.3	0.53	123	85	0.56	112.7
35	88.2	0.54	133	91.2	0.57	36	94.1	0.61	51	91.2	0.57	73.3
36	93.1	0.8	6	92.2	0.58	142	93.1	0.71	94	92.8	0.7	80.7
37	86.3	0.53	166	94.1	0.68	51	93.1	0.66	170	91.2	0.62	129
Average	83.2	0.6	90.5	84.0	0.61	100.0	84.5	0.63	97.8	**83.9**	**0.61**	**96.3**
ALL	61.9	0.62	190	62.8	0.63	142	58.0	0.58	159	**60.9**	**0.61**	**163.6**

A*– Classification Accuracy F1#– F1 Score F+– Optimal Number of Features

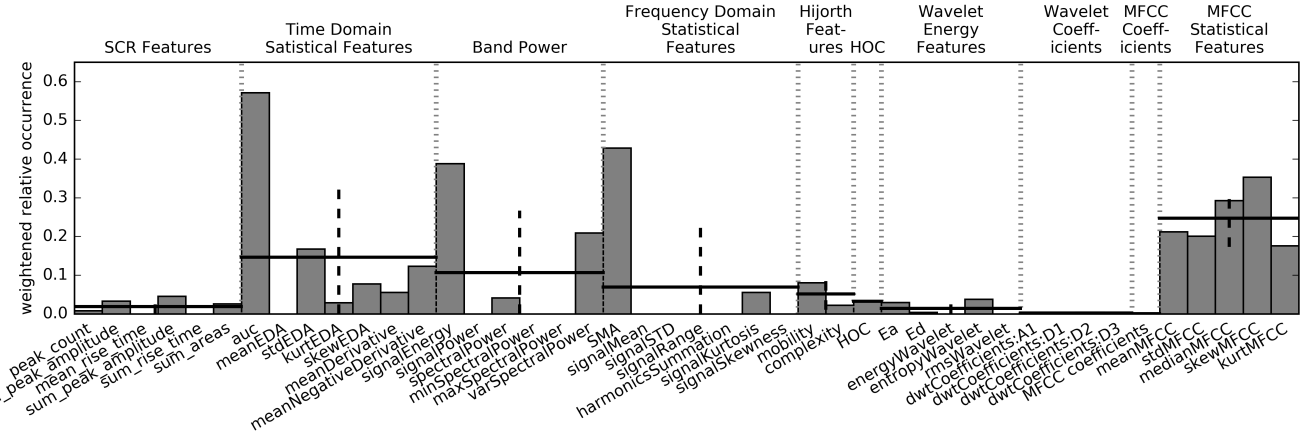

Fig. 5. Weighted relative frequency of each EDA feature type for arousal recognition. Black horizontal bar represents the mean for the group, and the vertical dashed black line represents the standard deviation for each group.

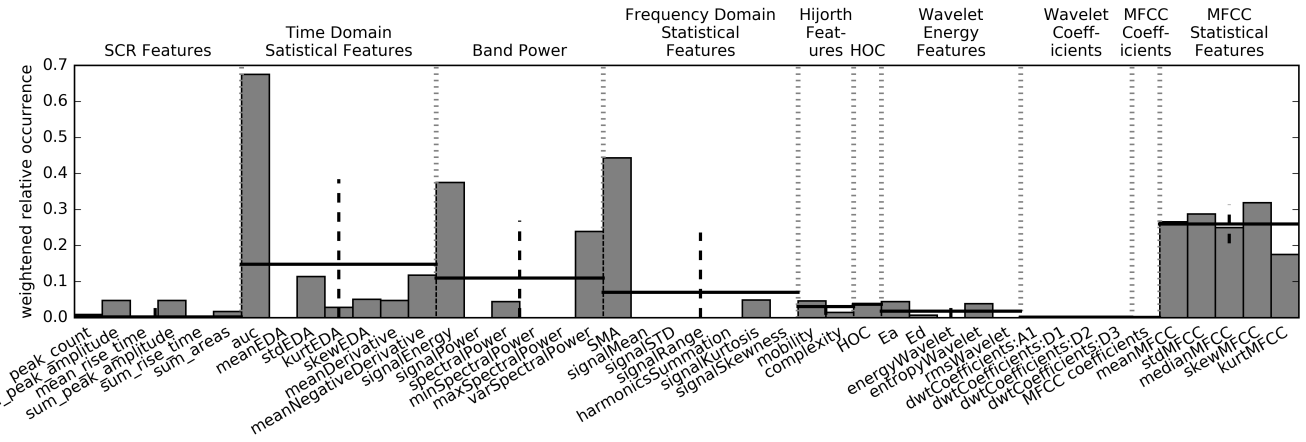

Fig. 6. Weighted relative frequency of each EDA feature type for valence recognition. Black horizontal bar represents the mean for the group, and the vertical dashed black line represents the standard deviation for each group.

3) Finally, while averaging, these relative frequencies were weighted across the FS methods by multiplying with the achieved classification accuracy.

Employing the statistics above, the relative frequency of the feature types varies between 0 and 1 and the most significant features score higher than the non-significant features for emotion recognition. Figure 5 and Figure 6 show the weighted relative frequency of each feature type for arousal and valence recognition, respectively. The most frequently selected features group for arousal recognition is the MFCC statistical features. Statistical features related to SCR of the EDA signal in the time domain and the band power related features in the frequency domain had the second best performance among all the feature groups. The single best performing feature among all the employed features is the AUC feature followed by SMA and the signal energy feature. The standard deviation or variance of the SCR signal along with the derivative features of the SCR signal also yield a better performance compared to other time domain statistical features. In general, the statistical features of the SCR signal perform better than the SCR related features in the time domain. It is worth noting that the time domain statistical features, band power and the frequency domain statistical features show a higher variance than the other feature groups. It indicates that particular features in these groups are more informative than others. These feature types are AUC, signal energy and SMA in the time domain statistical features, band power and the frequency domain statistical features, respectively. The variance among the MFCC statistical features is lower in comparison to the three other feature types. It means that all the features in this type are significant. The least frequently selected features are the coefficients of the wavelet and the coefficients of the MFCC, as they have low weighted relative occurrence scores.

We found that the feature usage for valence recognition follows the same trend as that of arousal recognition. The most frequently selected feature group for valence recognition is the MFCC statistical features, also followed by the time domain and the band power related features in the frequency domain. The single best performing feature among all the employed features is also the AUC feature followed by SMA and the signal energy feature.

5 Discussion

We did not find any significant difference in the performance of the three FS methods employed in this study, but all of them yielded a high average classification accuracy and F1-scores for both arousal and valence recognition. The results obtained in Table 3 and Table 4 indicate that a high number of EDA features ($\sim$95 for arousal recognition and $\sim$96 for valence recognition) are required to obtain the optimal accuracy. This is a significant finding and have not been reported by any earlier study. The similarity between the optimal number of features for arousal and valence recognition suggests that the computational complexity of the recognition for these two dimensions of emotions is quite similar.

While EDA is generally more related to emotional arousal [4], we found similar recognition performances for both arousal and valence. In this regard, the results reported by [38] (e.g. F-1 scores of 0.528 for valence and 0.541 for arousal for all videos, using a different set of signals and features), although showing a slightly better performance for arousal, also show that the classification performance for the two variables does not diverge considerably, which is aligned with our results. We found that there was a significant and quite high correlation between the annotated arousal and valence scores of AMIGOS dataset, $r = 0.56, p<.001$, which may help to explain why the recognition performance for both arousal and valence is similar.

Regarding the significance of specific features, it is worth noting that SCR features that have been widely used in the literature yielded low weighted relative occurrence scores in our study. Our results revealed that, among all SCR features, the amplitude of the SCR peaks is the most significant one for recognition of both arousal and valence from EDA. We also demonstrated that the commonly used rise time feature does not play an important role in this regard. Whereas AUC of the EDA signal has not been usually exploited in the literature, as a single feature it yields a high performance

for emotion recognition from EDA signals. This is also true for SMA and signal energy features of the EDA signal. Furthermore, the most significant finding of our study is the performance of the statistical features related to MFCC, which outperformed all other feature types across all the three domains. Taken together, these results suggest that EDA features with the highest potential for emotion classification have been either never used in previous studies (MFCC feature) or have been used with limited exploitation (AUC and SMA features).

Superior subject-dependent classification accuracies, using a lower number of features, were obtained in comparison to the subject-independent classification. The big difference in such classification results is not surprising. It is known that different individuals usually have different physiological responses to the same stimuli [54]. In addition, non-emotional individual contexts vary in a complex manner among different subjects [11]. If the subject is known in advance to the system or if the system can undergo a learning phase for each subject prior to the classification, then the emotion classification can be done in a user-dependent way. We believe that this is one of the biggest challenges of real-time emotion recognition and it goes beyond the scope of this article.

6 Conclusion

In this paper, we reviewed and implemented 40 different EDA features for emotion recognition suggested by 25 studies. We analyzed the significance and the suitability of different EDA feature across time, frequency and time-frequency domains using three FS methods, JMI, CMIM and DISR, and employing machine learning techniques on the publicly available AMIGOS dataset. All the three FS methods indicated use of $\sim$95 features on an average for arousal recognition and of $\sim$96 features on an average for valence recognition. The results reported an average accuracy of 85.75% (F1-score: 0.63) for arousal recognition and an average accuracy of 83.9% (F1-score: 0.61) for valence recognition. Also, the subject-dependent classification results were significantly higher than the subject-independent classification for both arousal and valence recognition. Statistical MFCC features along with the AUC and SMA features outperformed the commonly used SCR features of the EDA signal.

Our research not only describes the significance of a more exhaustive set of features than previous studies, but also points out the utility and informative properties for emotion classification of a new set of features from EDA signal that have been largely neglected in previous research, such as MFCC features. By doing so, it opens venues for the future development of new emotion recognition systems based on EDA with higher accuracy and minimizing its computational cost, which is key for the development of emotion detection applications that may work in real time.

One limitation of the current study is that we only relied in a dimensional model of emotions (i.e. the conceptualization of emotions in terms of valence and arousal), since the annotations of the AMIGOS dataset were provided in terms of arousal and valence. It may also be interesting to examine the predictive power of EDA features for classifying discrete emotions like joy, sadness, fear, or surprise, and thus this is an aspect that should be addressed in future studies. Another limitation of the present work is that we tested only a specific context for emotion elicitation (i.e. the context used in the AMIGOS study, that is, watching emotional videos). It is possible that different emotional contexts (e.g. a stressing job interview) may produce different patterns in EDA signal that might be better captured by focusing on different features of the signal.

Our results also stress the need to account for individual differences in subjects with different psychophysiological profiles, which tend to have different physiological responses for the same stimuli [54]. Failure to address this individual variability can negatively affect the classification performance of emotional state, which is shown by the results of subject-independent classification in our results. To tackle this issue, a general model for emotion classification could be trained with a sufficiently higher number of subjects, and then such a model could be fine tuned with the baseline values of the new users.

Acknowledgments

This research work is supported by the Industrial Doctorate program (Ref. ID.: 2014-DI-022) of AGAUR, Govt. of Catalonia, is partly funded by IIITA (Ref. ID.: URV.B05.01.01 N-007135) and is partly supported by the Agència per a la Competivitat de l´Empresa, ACCIÓ. J. Shukla is partly supported through Infosys Center for Artificial Intelligence at IIIT-Delhi, India.

References

[1] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, Oct 2001.

[2] J. Shukla, M. Barreda-Ángeles, J. Oliver, and D. Puig, "Muderi: Multimodal database for emotion recognition among intellectually disabled individuals," in *Social Robotics*, A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He, Eds. Cham: Springer International Publishing, 2016, pp. 264–273.

[3] G. Perugia, D. Rodrguez-Martn, M. D. Boladeras, A. C. Mallofr, E. Barakova, and M. Rauterberg, "Electrodermal activity: Explorations in the psychophysiology of engagement with social robots in dementia," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2017, pp. 1248–1254.

[4] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system," in *Handbook of psychophysiology, 3rd ed*, J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, Eds. New York, NY, US: Cambridge University Press, 2007, pp. 159–181.

[5] W. Boucsein, *Electrodermal Activity*. Boston, MA: Springer US, 2012.

[6] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxeda: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, April 2016.

[7] A. Drachen, L. E. Nacke, G. Yannakakis, and A. L. Pedersen, "Correlation between heart rate, electrodermal activity and player experience in first-person shooter games," in *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, ser. Sandbox '10. New York, NY, USA: ACM, 2010, pp. 49–54.

[8] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, June 2005.

[9] M. Swangnetr and D. B. Kaber, "Emotional state classification in patient-robot interaction using wavelet analysis and statistics-based feature selection," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 1, pp. 63–75, Jan 2013.

[10] J. Wang and Y. Gong, "Recognition of multiple drivers's emotional state," in *2008 19th International Conference on Pattern Recognition*, Dec 2008, pp. 1–4.

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

[12] M. D. van der Zwaag, J. H. Janssen, and J. H. D. M. 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, Jan 2013.

[13] R. Piacentini, "Emotions at fingertips, revealing individual features in galvanic skin response signals," Ph.D. dissertation, Universit degli studi di Roma, Roma, 2004.

[14] R. G. OConnell, M. A. Bellgrove, P. M. Dockree, A. Lau, M. Fitzgerald, and I. H. Robertson, "Self-alert training: Volitional modulation of autonomic arousal improves sustained attention," *Neuropsychologia*, vol. 46, no. 5, pp. 1379 – 1390, 2008.

[15] J. Zhai, A. B. Barreto, C. Chin, and C. Li, "Realization of stress detection using psychophysiological signals for improvement of human-computer interactions," in *Proceedings. IEEE SoutheastCon, 2005.*, April 2005, pp. 415–420.

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

[17] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from eeg," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, July 2014.

[18] R. Horlings, D. Datcu, and L. J. M. Rothkrantz, "Emotion recognition using brain activity," in *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, ser. CompSysTech '08. New York, NY, USA: ACM, 2008, pp. 6:II.1–6:1.

[19] Y. Hller, J. Bergmann, A. Thomschewski, M. Kronbichler, P. Hller, J. S. Crone, E. V. Schmid, K. Butz, R. Nardone, and E. Trinka, "Comparison of eeg-features and classification methods for motor imagery in patients with disorders of consciousness," *PLOS ONE*, vol. 8, no. 11, pp. 1–15, 11 2013.

[20] Y. Shimomura, T. Yoda, K. Sugiura, A. Horiguchi, K. Iwanaga, and T. Katsuura, "Use of frequency domain analysis of skin conductance for evaluation of mental workload," *Journal of PHYSIOLOGICAL ANTHROPOLOGY*, vol. 27, no. 4, pp. 173–177, 2008.

[21] P. Ghaderyan and A. Abbasi, "An efficient automatic workload estimation method based on electrodermal activity using pattern classifier combinations," *International Journal of Psychophysiology*, vol. 110, pp. 91 – 101, 2016.

[22] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis ;using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, Jan 2012.

[23] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *Journal of Biomedical Informatics*, vol. 59, pp. 49 – 75, 2016.

[24] C. S. Lima, A. Tavares, J. H. Correia, M. J. Cardoso, and D. Barbosa, *Non-Stationary Biosignal Modelling*, 2010, exported from https://app.dimensions.ai on 2018/09/26. [Online]. Available: https://app.dimensions.ai/details/publication/pub.1041870291 and http://www.intechopen.com/download/pdf/9075

[25] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications." Springer-Verlag, 1995, pp. 281–300.

[26] J. Shukla, M. Barreda-Ángeles, J. Oliver, and D. Puig, "Efficient wavelet-based artifact removal for electrodermal activity in real-world applications," *Biomedical Signal Processing and Control*, vol. Under Review, 2017.

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

[28] J. Benesty, M. Sondhi, and Y. Huang, *Springer Handbook of Speech Processing*, ser. Springer Handbook of Speech Processing. Springer Berlin Heidelberg, 2007.

[29] M. Li and S. Narayanan, "Robust ecg biometrics by fusing temporal and cepstral information," in *2010 20th International Conference on Pattern Recognition*, Aug 2010, pp. 1326–1329.

[30] C. Kamath, "Teager energy based filter-bank cepstra in eeg classification for seizure detection using radial basis function neural network," *ISRN Biomedical Engineering*, vol. 2013, p. 9, 2013.

[31] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug 1980.

[32] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 625–629, Oct 2008.

[33] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, "Conditional likelihood maximisation: A unifying framework for information theoretic feature selection," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 27–66, Jan. 2012.

[34] H. H. Yang and J. Moody, "Data visualization and feature selection: New algorithms for nongaussian data," in *in Advances in Neural Information Processing Systems*. MIT Press, 1999, pp. 687–693.

[35] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, pp. 1531–1555, Dec. 2004.

[36] C. Freeman, D. Kuli, and O. Basir, "An evaluation of classifier-specific filter measure performance for feature selection," *Pattern Recognition*, vol. 48, no. 5, pp. 1812 – 1826, 2015.

[37] P. E. Meyer and G. Bontempi, "On the use of variable complementarity for feature selection in cancer classification," in *Applications of Evolutionary Computing*, F. Rothlauf, J. Branke, S. Cagnoni, E. Costa, C. Cotta, R. Drechsler, E. Lutton, P. Machado, J. H. Moore, J. Romero, G. D. Smith, G. Squillero, and H. Takagi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 91–102.

[38] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," 2017.

[39] D. Giakoumis, D. Tzovaras, K. Moustakas, and G. Hassapis, "Automatic recognition of boredom in video games using novel biosignal moment-based features," *IEEE Transactions on Affective Computing*, vol. 2, no. 3, pp. 119–133, July 2011.

[40] A. Greco, G. Valenza, L. Citi, and E. P. Scilingo, "Arousal and valence recognition of affective sounds based on electrodermal activity," *IEEE Sensors Journal*, vol. 17, no. 3, pp. 716–725, Feb 2017.

[41] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, Sept 2013, pp. 671–676.

[42] A. Barreto, J. Zhai, and M. Adjouadi, "Non-intrusive physiological monitoring for automated stress detection in human-computer interaction," in *Human–Computer Interaction*, M. Lew, N. Sebe, T. S. Huang, and E. M. Bakker, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 29–38.

[43] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and galvanic skin response signals," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, June 2013, pp. 209–214.

[44] S. A. Hosseini and M. A. Khalilzadeh, "Emotional stress recognition system using eeg and psychophysiological signals: Using new labelling process of eeg signals in emotional stress state," in *2010 International Conference on Biomedical Engineering and Computer Science*, April 2010, pp. 1–6.

[45] I. Leite, R. Henriques, C. Martinho, and A. Paiva, "Sensors in the wild: Exploring electrodermal activity in child-robot interaction," *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 41–48, 2013.

[46] D. Giakoumis, D. Tzovaras, and G. Hassapis, "Subject-dependent biosignal features for increased accuracy in psychological stress detection," *International Journal of Human-Computer Studies*, vol. 71, no. 4, pp. 425 – 439, 2013.

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

[48] S. Yakowitz, "Time series analysis of higher order crossings (b. kedem)," *SIAM Review*, vol. 36, no. 4, pp. 680–682, 1994.

[49] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, and A. Nowe, "A survey on filter techniques for feature selection in gene expression microarray analysis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 4, pp. 1106–1119, July 2012.

[50] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE Inter-*

national Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), June 2008, pp. 1322–1328.
[51] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human–robot interaction," *Pattern Analysis and Applications*, vol. 9, no. 1, pp. 58–69, 2006.
[52] S. A. Hosseini, M. A. Khalilzadeh, M. B. Naghibi-Sistani, and V. Niazmand, "Higher order spectra analysis of eeg signals in emotional stress states," in *2010 Second International Conference on Information Technology and Computer Science*, July 2010, pp. 60–63.
[53] M. Li and B. L. Lu, "Emotion classification based on gamma-band eeg," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sept 2009, pp. 1223–1226.
[54] R. Henriques and A. Paiva, "Learning effective models of emotions from physiological signals: The seven principles," in *Physiological Computing Systems*, H. P. da Silva, A. Holzinger, S. Fairclough, and D. Majoe, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 137–155.

Jainendra Shukla completed his Ph.D. from Universitat Rovira i Virgili (URV), Spain in 2018 and his doctoral thesis was Empowering Cognitive Stimulation Therapy with Socially Assistive Robotics and Emotion Recognition. Earlier, he obtained his M.Tech. degree in Robotics from Indian Institute of Information Technology, Allahabad (IIITA), India in 2012 and B.E. degree in Information Technology from University of Mumbai, India in 2009. His research interests include Socially Assistive Robotics (SAR) and emotion recognition using physiological signals.

Miguel Barreda-Ángeles received the BA degree in audiovisual communication from the University of Seville, 2005, the MA degree in television innovation from Pompeu Fabra University and Autonomous University of Barcelona, 2007, and a PhD in Social Communication from Pompeu Fabra University, 2014. He has worked as a researcher in the Perception and Cognition research group at Fundaciò Barcelona Media, and currently is a researcher within the Data Science & Big Data Analytics unit at Eurecat, Centre Tecnològic de Catalunya, Barcelona. His research interests include media psychology and the use of psychophysiological methods for analyzing user experience with media technologies.

Joan Oliver (M.Sc. in Industrial Engineering) is Director at Instituto de Robtica para la Dependencia and PhD student at Universitat Rovira i Virgili. He has 14 years of experience developing robots and robotic solutions for several markets. He has experience managing teams, projects and also in business development and entrepreneurship.

G. C. Nandi graduated from Indian Institute of Engineering, Science and Technology, Shibpur in 1984 and completed his post graduation from Jadavpur University, Calcutta in 1986. He obtained his Ph.D. degree from the Russian Academy of Sciences, Moscow in 1992. He was a visiting research scientist at the Chinese University of Hong Kong in 1997 and faculty at Carnegie Mellon University (CMU), USA in 2010-11. Currently, he is the senior most professor at the Indian Institute of Information Technology, Allahabad (IIITA), India. He is a senior member of the ACM and IEEE. He has published more than 100 papers in various refereed journals and international conferences. His research interest includes robotics, especially biped locomotion control and humanoid push recovery, artificial intelligence, soft computing, and computer controlled systems.

Domènec Puig received a degree in computer science and a doctorate from the Polytechnic University of Catalonia. In 1992, he joined the Department of Computer Engineering and Mathematics of the University Rovira i Virgili (URV), where he is Full Professor. Since 2006, he is the principal investigator of the Intelligent Robotics and Computer Vision research group of the URV. His research interests include image processing, pattern recognition, computer vision, machine learning, perception models, mobile robotics and socially assistive robotics. Prof. Domènec Puig has directed and/or participated in more than 50 R&D projects and published more than 150 papers.