DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices

Stamos Katsigiannis and Naeem Ramzan

***Abstract*—In this paper, we present DREAMER, a multi-modal database consisting of electroencephalogram (EEG) and electrocardiogram (ECG) signals recorded during affect elicitation by means of audio-visual stimuli. Signals from 23 participants were recorded along with the participants self-assessment of their affective state after each stimuli, in terms of valence, arousal, and dominance. All the signals were captured using portable, wearable, wireless, low-cost, and off-the-shelf equipment that has the potential to allow the use of affective computing methods in everyday applications. A baseline for participant-wise affect recognition using EEG and ECG-based features, as well as their fusion, was established through supervised classification experiments using support vector machines (SVMs). The self-assessment of the participants was evaluated through comparison with the self-assessments from another study using the same audio-visual stimuli. Classification results for valence, arousal, and dominance of the proposed database are comparable to the ones achieved for other databases that use nonportable, expensive, medical grade devices. These results indicate the prospects of using low-cost devices for affect recognition applications. The proposed database will be made publicly available in order to allow researchers to achieve a more thorough evaluation of the suitability of these capturing devices for affect recognition applications.**

***Index Terms*—affect, affect recognition, ECG, EEG, emotion, physiological signals, wireless devices.**

I. Introduction

AFFECT recognition is the process of understanding what type of affect (emotion) a person is expressing. The development of efficient and robust algorithms for automated recognition of human affect is a major challenge for the field of affective computing and may have great implications on the way people interact with computing devices. Human-computer interaction is one of the primary examples of computing disciplines that rely on knowledge about human emotions and on knowledge about how acoustic or optical stimuli can affect the emotional state and the affective expressions of humans. The introduction of automated affect recognition techniques in human-computer interaction applications could significantly increase the quality of the user experience and lead to more emotional-aware computer interfaces.

Another area that could significantly benefit from progress in affect recognition is the multimedia industry. Video and audio content is usually annotated in terms of the emotions that the content is expected to elicit through the use of genre types (e.g. a drama is expected to elicit sadness). Annotation is performed manually, usually by taking into consideration the opinion of industry experts, critics and audience. Algorithms for automated prediction of the expected human emotional response to the multimedia content would allow the mass annotation of multimedia content with keywords that would be more relevant to the emotions actually elicited by the content [1]. Moreover, taking into consideration that the main focus of multimedia content is to elicit specific emotions to the audience, it is evident that the introduction of algorithmic methods for emotion prediction and measurement could significantly benefit multimedia creators. Relations between stimuli and emotional state and expressions have been extensively studied by psychologists. Nevertheless, algorithmic affect recognition remains an arduous and difficult task since human emotion manifestation involves multiple types of responses, spanning from affective intonations and facial expressions to physiological responses that originate from the human nervous system [2].

The first step for achieving automated affect recognition is to conceptualise affect in a strict and clear manner. Psychological research has provided ways to conceptualise affect in the form of discrete categories of emotions or in terms of a small number of latent dimensions [3]. The use of discrete categories for describing emotions has been a long standing approach due to being based on the language used by humans. One of the most popular discrete categorisations of emotions is the use of the six basic emotions (happiness, sadness, anger, fear, surprise and disgust) proposed by Ekman *et al.* [4], whose use has also been supported by cross-cultural studies, indicating that regardless of culture, humans perceive these basic emotions in a similar manner [5]. Another popular discrete categorisation of emotions is Parrott's [6] tree structure of emotions. These categorisations of emotions seem very intuitive and match the way people categorise emotions in daily life. Nevertheless, the

Manuscript received September 6, 2016; revised February 20, 2017; accepted March 19, 2017. Date of publication March 27, 2017; date of current version January 3, 2018. This work was supported in part by the UWS VP Fund, in part by the Construction Scotland Innovation Centre, and in part by Wheatley Group/Lorretto Care. *(Corresponding author: Stamos Katsigiannis.)*

The authors are with the School of Engineering and Computing, University of the West of Scotland, Paisley PA1 2BE, U.K. (e-mail: Stamos.Katsigiannis@uws.ac.uk; Naeem.Ramzan@uws.ac.uk).

Digital Object Identifier 10.1109/JBHI.2017.2688239

use of discrete emotions is not suitable for describing the whole range of emotions that may appear in natural settings and thus a more pragmatic and context dependent manner for describing emotions is needed [3].

The dimensional description of emotion offers a suitable alternative to the discrete categorisation [7], [8]. Dimensional scales like the emotion wheel proposed by Plutchik [9] and the valence-arousal scale proposed by Russell [7] characterise emotion in terms of dimensions that correspond to the main aspects of emotions. In the widely used work by Russell [7], the valence dimension measures whether a human has negative or positive feelings, whereas the arousal dimension measures whether a human feels bored or excited. Each perceived emotional state can be depicted on a 2-dimensional plane with valence and arousal at each axis respectively. An expansion of this approach [10] proposed the use of a third dimension, called dominance, which measures whether the human feels without control or empowered. Taking into consideration that most of the variance in emotion relates to the arousal and valence dimensions, most works on affect recognition use only these two dimensions [1], [11]–[13]. Nevertheless, this categorisation of emotion is not intuitive for humans and as a result, subjects participating in emotion recognition studies should be trained on using the dimensional emotion labelling system. The vast set of different human emotional expressions poses a great challenge for affect recognition research.

The employment of pattern recognition approaches for affect recognition requires the acquisition of data regarding the affective state of a subject during the period that a specific emotion is expressed. Various modalities have been used for affect related data acquisition, spanning from facial expressions, peripheral physiological signals (e.g. electrocardiogram (ECG), electromyogram (EMG), electroencephalogram (EEG), galvanic skin response (GSR)), and speech intonations, to imaging modalities such as fMRI [2] and thermal infrared imaging [14]. While initial approaches made use of a single modality, more recent approaches opt to use multi-modal data as it has been shown to provide higher recognition accuracy [1]–[3], [12]. Another important aspect of affect recognition studies is the selection of suitable stimuli that will be used for eliciting emotions in human subjects. Stimuli may include audio-only stimuli, video-only, or a combination of audio and video in the form of film clips or music video clips. Several databases containing stimuli media as well as emotional expressions through different modalities have been created in recent years. Databases like the Gunes *et al.* [15], Fanelli *et al.* [16], Grimm *et al.* [17], DEAP [12], MAHNOB-HCI [1] and DECAF [2] have been used for affect recognition using pattern recognition methods on signals coming from multiple modalities, achieving different levels of success. One of the common characteristics of these databases is that specialised, non-portable and expensive equipment was utilised in order to capture the bio-signals included. Although the use of such devices enables the study of EEG and physiological signals in relation to affect, it restricts the application of the proposed algorithms into confined and controlled spaces that can accommodate the use of such devices. As a result, the application of such algorithms cannot be expanded into casual and everyday scenarios that could benefit from affective computing.

In order to address this limitation of the available databases for emotion analysis, in this work we propose DREAMER, a database consisting of recordings of EEG and ECG signals captured while audio-visual stimuli was presented to the participants of this study in order to elicit specific emotions. Portable, wearable, wireless, low-cost and off-the-shelf devices were used for both the EEG and ECG capture allowing the evaluation of affect recognition algorithms on signals that can be conveniently captured in everyday scenarios, providing the means to integrate affect computing methods into a wide variety of tasks. Audio-visual stimuli in the form of clips taken from known films was used for eliciting emotional responses to the participants who then proceeded to rate the felt emotions according to three rating scales. After assessing the quality of the participants ratings, a baseline for this database was then established by evaluating EEG and ECG -based features through participant-wise supervised classification experiments. The motivation of this work is threefold: 1) To provide a dataset for affect recognition research, consisting of EEG and ECG signals captured using low-cost portable devices, 2) To compare the affect recognition -related performance of these devices against "medical grade" devices and establish whether they constitute a viable alternative, and 3) To demonstrate the feasibility of integrating affective computing methods to everyday applications through the use of portable/wearable equipment.

The rest of this paper is organised in six sections. Section II provides an overview of the most recent works in the field of affect recognition. Section III describes in detail the experimental procedure, while the proposed data processing, feature extraction and classification methods are described in Section IV. Statistics and results of the classification procedure are presented in Section V and discussed in Section VI. Finally, conclusions are drawn in Section VII.

II. Background

Given the importance and the interest in the field of affect recognition, there have been multiple proposed approaches in the domain of affect recognition from physiological signals and/or other modalities. Zeng *et al.* [3] provide an extensive survey on affect recognition methods, proposed till 2009, using video, audio or both audio and video as stimuli for affect elicitation. Furthermore, in another more recent survey, Gunes *et al.* [18] present an extensive review on continuous affect detection. Nicolaou *et al.* [19] proposed the use of features based on facial expression, shoulder gesture and audio cues along with Support Vector Machines for Regression (SVRs) and Bidirectional Long Short-Term Memory Neural Networks (BLSTM-NNs) for classification. Results showed that arousal was better predicted using audio cues, while multi-modal approaches were more successful for valence. Soleymani *et al.* [1] evaluated the use of features based on eye gaze data, EEG and peripheral physiological signals along with a Support Vector Machine (SVM) classifier for affect recognition on their proposed MAHNOB-HCI database. Results showed that the fusion of the best two modalities (EEG

TABLE I
MEAN RATING AND STANDARD DEVIATION ACROSS ALL PARTICIPANTS FOR EACH STIMULI FILM CLIP AND FOR EACH RATING SCALE

ID	Film clip	Target emotion	Valence	Arousal	Dominance
1	Searching for Bobby Fischer	calmness	3.17 ± 0.72	2.26 ± 0.75	2.09 ± 0.73
2	D.O.A.	surprise	3.04 ± 0.88	3.00 ± 1.00	2.70 ± 0.88
3	The Hangover	amusement	4.57 ± 0.73	3.83 ± 0.83	3.83 ± 0.72
4	The Ring	fear	2.04 ± 1.02	4.26 ± 0.69	4.13 ± 0.87
5	300	excitement	3.22 ± 1.17	3.70 ± 0.70	3.52 ± 0.95
6	National Lampoon's VanWilder	disgust	2.70 ± 1.55	3.83 ± 0.83	4.04 ± 0.98
7	Wall-E	happiness	4.52 ± 0.59	3.17 ± 0.98	3.57 ± 0.99
8	Crash	anger	1.35 ± 0.65	3.96 ± 0.77	4.35 ± 0.65
9	My Girl	sadness	1.39 ± 0.66	3.00 ± 1.09	3.48 ± 0.95
10	The Fly	disgust	2.17 ± 1.15	3.30 ± 1.02	3.61 ± 0.89
11	Pride and Prejudice	calmness	3.96 ± 0.64	1.96 ± 0.82	2.61 ± 0.89
12	Modern Times	amusement	3.96 ± 0.56	2.61 ± 0.89	2.70 ± 0.82
13	Remember the Titans	happiness	4.39 ± 0.66	3.70 ± 0.97	3.74 ± 0.96
14	Gentlemans Agreement	anger	2.35 ± 0.65	2.22 ± 0.85	2.39 ± 0.72
15	Psycho	fear	2.48 ± 0.85	3.09 ± 1.00	3.22 ± 0.9
16	The Bourne Identity	excitement	3.65 ± 0.65	3.35 ± 1.07	3.26 ± 1.14
17	The Shawshank Redemption	sadness	1.52 ± 0.59	3.00 ± 0.74	3.96 ± 0.77
18	The Departed	surprise	2.65 ± 0.78	3.91 ± 0.85	3.57 ± 1.04

and eye gaze) provided the highest classification accuracy for both arousal and valence, while the accuracy achieved for the features based on peripheral physiological responses had high variance between the participants in the study, leading to reduced overall performance. EEG and peripheral physiological signals were also used by Koelstra *et al.* [12] on their proposed DEAP database. Their experimental evaluation using the Naive Bayes classifier showed that EEG-based features achieved higher accuracy for arousal, while the features based on peripheral physiological signals performed better for valence. Nevertheless, both modalities were outperformed by features based on the audio and visual content of the music videos used for affect elicitation. AlZoubi *et al.* [20] utilized a tutoring software for affect elicitation and evaluated the use of features based on Electromyogram (EMG), Galvanic Skin Response (GSR) and ECG physiological data, along nine classifiers for affect detection. Results showed that the k-Nearest Neighbour (k-NN) and the Linear Bayes Normal Classifier (LBNC) provided the best accuracy in affect recognition and that the use of features based on a single modality or the fusion of all modalities yielded better results in general compared to the fusion of two of the proposed modalities. More recently, Abadi *et al.* [2] presented DECAF, a multi-modal data set for decoding user physiological responses to affective multimedia content. Instead of the commonly used EEG modality, Abadi *et al.* proposed the use of Magnetoencephalogram (MEG)-based features, along with other peripheral physiology signals (horizontal Electrooculogram (hEOG), trapezius-Electromyogram (tEMG) and facial data, ECG), for affect recognition. Their experimental evaluation showed that MEG signals effectively encoded arousal and dominance, while the peripheral physiology signals were more successful for valence. Valence was also successfully encoded by facial features, while the fusion of all modalities provided slightly lower results compared to different combinations of features. Facial expression features were also recently used for continuous affect detection along EEG-based features and Long-Short-Term-Memory recurrent Neural Networks (LSTM-RNN) and Continuous Conditional Random Fields (CCRF) by Soleymani *et al.* [11]. Their experimental evaluation showed that features based on facial expressions were superior to both the EEG-based features and the fusion of facial expression-based and EEG-based features. Nevertheless, their analysis showed that EEG signals contained complementary information in presence of facial expressions.

Specialised and non-portable equipment was used for capturing the EEG and the other physiological signals in all the aforementioned works. Systems like the commonly used Biosemi Active II (DEAP [12], MAHNOB-HCI [1]), the ELEKTA Biomag (DECAF [2]) and the BIOPAC MP150 (AlZoubi *et al.* [20]) provide enhanced capturing capabilities and increased signal quality but are costly, non-portable and non-wearable and as a result they are not suitable for casual everyday applications.

III. EXPERIMENT PROTOCOL

A. Data Acquisition

Audio and visual stimuli in the form of film clips was employed in order to elicit emotional reactions to the participants of this study and record EEG and ECG data. A dataset consisting of 18 film clips selected and evaluated by Gabert-Quillen *et al.* [21] was utilised for eliciting emotions. These film clips contain cut out scenes from different films that have been shown to evoke a wide range of emotions. From these 18 film clips, two of each targeted one of the following nine emotions: amusement, excitement, happiness, calmness, anger, disgust, fear, sadness and surprise. Details about the emotion targeted by each film clip are shown in Table I. The length of the film clips was between 65 to 393 s ($M = 199$ s), which is considered as sufficient since, according to psychologists, video stimuli between 1 to 10 min is capable of eliciting single emotions [11], [22]. Nevertheless, the emotional state of a person may change over time, especially when video stimuli of larger length is used. To avoid contaminating data recordings with multiple emotions, only the

Fig. 1. The Emotiv EPOC wireless EEG headset and the wireless SHIMMER ECG sensor.

recordings captured during the last 60 s of each film clip were used for further analysis.

The experiments were performed in an isolated environment with controlled illumination in order to avoid external influences. An electric curtain was used in order to completely darken the room and the video clips were presented on a 45" TV-monitor using the embedded speakers for audio playback. EEG was recorded at a sampling rate of 128 Hz using an Emotiv EPOC system [23], [24] that uses 16 gold-plated contact-sensors that are fixed to flexible plastic arms of a wireless headset and are placed against the head in locations aligned with the following locations according to the International 10-20 system: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, M1 and M2 [23]. The mastoid sensor at M1 acted as a ground reference point for comparing the voltage of all other sensors, while the mastoid sensor at M2 was a feed-forward reference for reducing external electrical interference. As a result, the signal from the other 14 contact-sensors was recorded and used for the later feature extraction step. ECG was recorded at 256 Hz using a SHIMMER$^{\mathrm{TM}}$ [25] wireless sensor which is able to produce RA→ LL and LA→ LL vectors. In this work, only the RA→ LL vector was later used for feature extraction. Fig. 1 shows the Emotiv EPOC headset and the SHIMMER ECG sensor.

For the evaluation of the emotions elicited by the film clips and the recording of physiological data, 25 healthy volunteers aged between 22 and 33 years old ($M = 26.6$, $SD = 2.7$) participated in the study. Before participating in the experiment, subjects were first asked to read an information sheet which provided details about the experimental procedure and explained the rating scales used for affect assessment and then proceed to sign a Consent Form. As mentioned in Section I, the categorisation of emotion using the valence/arousal/dominance scale is not intuitive for humans. To ensure that the participants would provide correct ratings, the rating scale was thoroughly explained both verbally and through examples. The session then started only if the participants expressed a sufficient understanding of the rating scales. A member of the team conducting the experiment was also available, in order to answer any questions posed by the participants. Each session lasted approximately one hour, during which the subjects were asked to minimise their movement. Moreover, they were asked to provide an emotional assessment derived by what emotions they actually felt and not what they thought that the video clip was intended to elicit.

Next, the experiment started by showing a neutral film clip, i.e. a video clip considered to have no valence in order to establish the baseline signals. This neutral video clip was the one used in the work of Gabert-Quillen *et al.* [21]. The use of neutral signals as baseline signals have been proven to allow the removal of daily dependencies when recording physiological data [12], [26]. As a result, the neutral clip was shown before each film clip in order to help the subject return to a neutral emotional state, in addition to establishing the baseline signal. After viewing each film clip, the subjects used a graphical user interface presented to them in order to evaluate their emotion by reporting the felt arousal (ranging from uninterested/bored to excited/alert), valence (ranging from unpleasant/stressed to happy/elated) and dominance (ranging from helpless to empowered) on five point scales. Self-assessment manikins (SAM) [27] were used in order to facilitate the subject's assessments of arousal, valence and dominance. Finally, the participants data, along with the recordings of the physiological data and the emotion assessments were stored in a data structure. The experimental protocol was implemented using the MATLAB environment [28] and its details are summarised in Table II.

TABLE II
EXPERIMENT SUMMARY

Audio-visual stimuli	
Number of videos	18
Video content	Audio-Video
Video duration	65–393 s ($M = 199$ s)
Experiment information	
Number of participants	25 (23)
Number of males	14 (14)
Number of females	11 (9)
Age of participants	22–33 ($M = 26.6$, $SD = 2.7$)
Rating scales	Arousal, Valence, Dominance
Rating values	1–5
Recorded signals	14-channel 128 Hz EEG, 256 Hz ECG

Furthermore, while the experimenters tried to carefully follow the experimental protocol and setup, problems arose with the recordings from two subjects due to technical problems, resulting to incomplete data. As a result, recordings from 23 out of the 25 volunteers were subsequently used in this work due to the unsuitability of the data from the two aforementioned subjects. From the final 23 subjects used in this study, 14 were male and 9 female.

This study, including the acquisition and publication of anonymised data, was approved by the University of the West of Scotland University Ethics Committee (UWS UEC).

B. Evaluation of Participants Self-Assessment

After collecting the subjects assessments, the data were analysed in order to detect any abnormalities or unexplained variations. While the evaluation of the emotions elicited by each video may have small variations for each participant, the overall assessment should not present extreme variations. In order to obtain a measure of the agreement between the participants of this study, the coefficient of variation (CV) for the participants assessments for all the film clips was computed. CV is defined as the ratio of the standard deviation and the mean and is a standardised measure of variability. A CV equal to zero denotes no variability across the samples (i.e. all samples are the

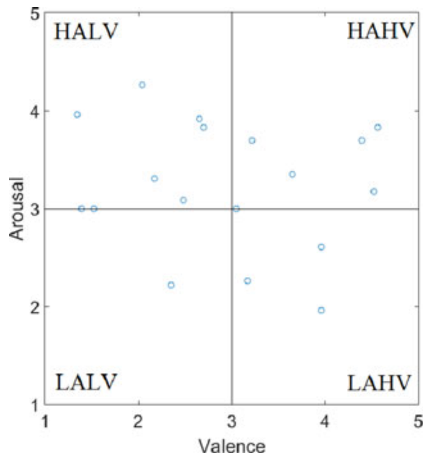

Fig. 2. Mean location of film clips in two-dimensional affective space as rated by the participants of this study.

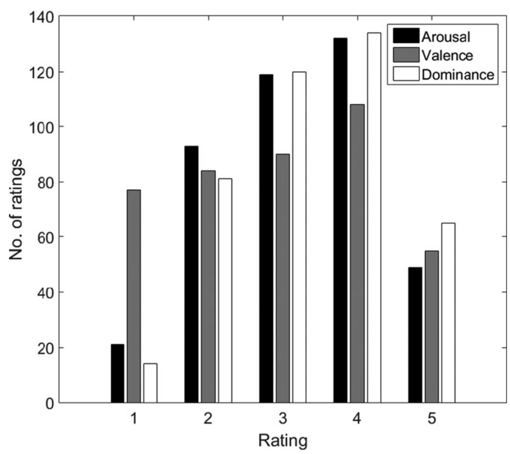

Fig. 3. Number of ratings by the participants at each scale of arousal, valence and dominance.

same) while higher *CV*s denote more variability. The mean *CV* between the participants assessments for arousal was 0.29 ± 0.07, for valence 0.31 ± 0.15 and for dominance 0.27 ± 0.06, showing that there was low variability between their assessments. Furthermore, the mean standard deviation for the assessment of each stimuli film clip was 0.81 for valence, 0.88 for arousal and 0.88 for dominance. The low mean standard deviations further support the claim of low variability between the participants assessments.

The mean assessments by all the participants in the study for each film clip in terms of valence and arousal are shown in Fig. 2 in the form of a scatter plot, while Table I shows the mean assessment across all participants for each stimuli film clip and for all rating scales. The assessments span the whole range of possible ratings leading to a balanced dataset, as can be seen on Fig. 3. The only misrepresented classes are the minimum arousal and dominance (for a rating equal to 1). Depending on the corresponding valence and arousal values, the valence-arousal space can be divided into four quadrants, as shown in Fig. 2. The four quadrants refer to each of the following possible combinations of valence/arousal states: low arousal-low valence (LALV), low arousal-high valence (LAHV), high arousal-low valence (HALV), and high arousal-high valence (HAHV). It is easily derived from Fig. 2 that the self-assessments of the subjects participating in this study are consistent with the emotions that the used film clips were intended to elicit [21]. A comparison between the subjective classification into the four valence-arousal quandrants from the subjects participating in this study compared to the subjective classification reported by Gabert-Quillen *et al.* [21] is shown in Table III. Furthermore, the Spearman's rank correlation coefficient between the subjective ratings by the participants of this study and the ratings reported by Gabert-Quillen *et al.* [21] was computed in order to evaluate their agreement and confirm the integrity of the collected ratings. The correlation analysis resulted in a Spearman's $\rho = 0.719$ for valence and $\rho = 0.870$ for arousal, indicating a strong correlation between the ratings from the two studies. It must be noted that the Gabert-Quillen *et al.* study did not collect ratings for dominance, thus a correlation with this study cannot be evaluated.

TABLE III
NUMBER OF FILM CLIPS SUBJECTIVELY CLASSIFIED ON AVERAGE INTO EACH CATEGORY

	LALV	LAHV	HALV	HAHV
This study	3	4	6	5
Gabert-Quillen *et al.* **[21]**	4	6	5	3

TABLE IV
CORRELATIONS BETWEEN THE RATINGS OF ALL PARTICIPANTS FOR THE SCALES OF VALENCE, AROUSAL AND DOMINANCE FOR ALL STIMULI

Scale	Valence	Arousal	Dominance
Valence	1	-0.02	-0.11^{*}
Arousal		1	0.70^{**}
Dominance			1

* *indicates a significant correlation according to the Student's t-test with* $p < 0.05$, *while* ** *denotes a significant correlation with* $p < 0.001$.

The ratings for the different scales are not completely independent since the participants may become biased to specific ratings due to lack of complete understanding of the rating scales or due to unwanted effects of habituation or fatigue due to the amount of time needed for the experiment [12]. The Spearman's rank correlation coefficient was computed between the ratings of all participants for the three rating scales in order to explore possible correlation. As shown in Table IV, a significant ($p < 0.001$) strong positive correlation ($\rho = 0.70$) was observed between arousal and dominance. Without implying any causality due to the correlation, it seems that when the participants felt excited about the stimuli (high arousal), they also felt empowered (high dominance), while when they felt bored and uninterested (low arousal) they also felt helpless and without control (low dominance). Furthermore, a non-significant ($p > 0.05$) very weak negative correlation ($\rho = -0.02$) was observed between valence and arousal, indicating that the participants were able to distinguish these two concepts. A significant

($p < 0.05$) but very weak negative correlation ($\rho = -0.11$) was also observed between valence and dominance.

IV. Data Analysis

A. Pre-processing

The captured EEG signals are contaminated with noise and artefacts that were detected but did not originate from the brain. Signals caused from cardiac activity, eye movement and muscular activity, as well as power line noise are also captured by the EEG device and thus downgrade the quality of the recorded data, making the use of denoising methods a necessity. On the other hand, ECG signals are less susceptible to interferences due to their higher voltage amplitudes and thus require no further processing.

In the EEG signals, most ocular artifacts (eye blinking, eye movement, cardiac interferences, etc.) are dominant below 4 Hz, muscle movements produce artefacts above 30 Hz [29] and power line noise usually lies at 50 or 60 Hz, while the frequency bands that contain information relative to the affect recognition task lie in the range of 4–30 Hz. This frequency range is commonly divided in the theta, alpha, and beta bands. Three separate bandpass Hamming sinc linear phase FIR filters are applied using the EEGLAB [30] toolbox in order to extract only the frequencies inside the ranges of interest. Artefacts that may have been introduced at the beginning or the end of the EEG data in the form of DC offsets are handled by padding the data by a DC constant at the beginning and at the end, before re-sampling using EEGLAB. Finally, the filtered EEG data are shifted by the filter's group delay. Furthermore, 50 and 60 Hz interference is not suppressed using notch filters, since the Emotiv EPOC headset comes with built-in digital notch filters at these frequencies [23].

The filtering process is not able to remove all the artefacts from the EEG signals and thus further processing is needed. The Artefact Subspace Reconstruction (ASR) method, proposed by Kothe [31], is used for artefact removal. The ASR method consists of a sliding window Principal Component Analysis (PCA), which statistically interpolates any high-variance signal components exceeding a threshold relative to the covariance of an automatically detected data section that is relatively free of artefacts. The final step for preparing the EEG data for further analysis is the application of the Common Average Reference (CAR) method, as recommended by Cohen [32], which computes the average value over all electrodes and subtracts it from each sample of each electrode. Removal of bad channels is performed prior to the CAR method in order to avoid introducing noise to all channels due to a possible bad channel.

B. Feature extraction

Due to the variable duration of the film clips used as stimuli and in order to allow time for a specific emotion to become dominant, the signal recordings corresponding to the last 60 seconds of each film clip were used for further feature extraction and analysis. Table V shows the features that are then computed from each modality, while an outline of the feature extraction procedure is provided on Fig. 4.

TABLE V
Extracted Features From Each Film Clip

Modality	Extracted features
EEG	theta, alpha and beta power spectral density (PSD) for each electrode
ECG	mean, median, standard deviation, min, max, and range from each part of the PQRST complexes, difference between consecutive RR intervals (RMSSD), PSD for Low Frequency (LF), PSD for High Frequency (HF), LF to HF ratio, total power

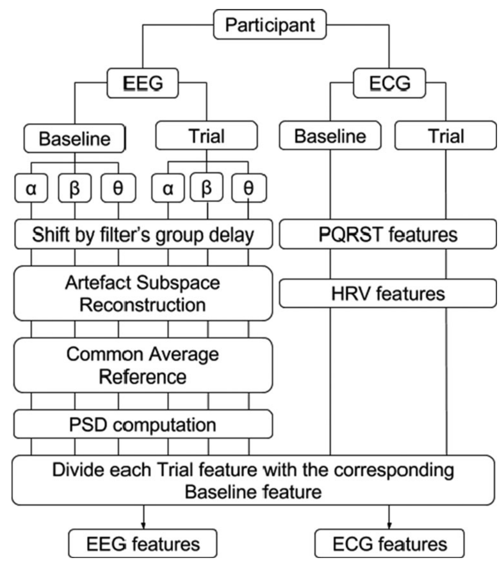

Fig. 4. Outline of the feature extraction procedure.

1) EEG-based features: It is well documented that the power spectral densities (PSDs) of EEG signals in different bands are correlated with the affective state of a human [33]. Soleymani *et al.* [11] showed that the higher frequency components of EEG signals carry more important information regarding positive emotions compared to negative ones (high and low valence respectively), while a correlation between increased beta power and positive emotional self-induction has also been reported [34]. Koelstra *et al.* [12] also found strong correlations between valence and EEG signals at all frequency bands. Furthermore, their study also found negative correlations between arousal and the theta, alpha and gamma bands of EEG signals, while prior studies [35], [36] have reported an inverse relationship between the power of the alpha band and the general arousal level.

After the preprocessing step, the captured EEG signals were separated into the theta (4 Hz–8 Hz), alpha (8 Hz–13 Hz), and beta (13 Hz–20 Hz) frequency bands. Welch's overlapped segment averaging estimator is then used to estimate the PSD of each EEG band, using a 256 samples window with an overlap of 128 samples. The logarithms of the PSD from each of the aforementioned bands are extracted from the signal of each of the 14 electrodes in order to be used as features, as also proposed in [1], [11], [12], [37], leading to a total of 42 features (3 for each of the 14 electrodes). Finally, all the features are concatenated into the final feature vector F_{EEG} as follows: Let

$F_{i\theta}$, $F_{i\alpha}$, and $F_{i\beta}$ be the logarithm of the PSD for the signal of the i-th electrode, $i = 1, 2, ..., 14$, for the theta, alpha and beta bands respectively. The final feature vector is defined as $F_{EEG} = [\, F_{1\theta} \quad F_{1\alpha} \quad F_{1\beta} \quad ... \quad F_{14\theta} \quad F_{14\alpha} \quad F_{14\beta} \,]$.

2) ECG-based features: Many studies have shown that features extracted from ECG signals correlate with changes in the affective state of a person [1], [2], [12], with the most consistently associated features usually being heart rate (HR) and heart rate variability (HRV) specific parameters in the time and frequency domain respectively. For example, heart rate variability may decrease with fear, sadness and happiness [38], while when a stimuli induces pleasantness, then the peak heart rate response may increase [39]. Moreover, spectral features derived from HRV have also been shown to correlate with affective state [40]. Taking into consideration the effect of the emotional state on ECG-based features, features derived from the HR and HRV of the recorded ECG signals are computed. The Pan-Tompkins QRS detection algorithm [41] is first used in order to detect the locations of QRS complexes within the ECG signal and accurately detect R-peaks. Statistical features, such as the mean, median, standard deviation, min, max, and range, are extracted from each part of the PQRST complexes of the ECG data using the Augsburg Biosignal Toolbox (AuBT) [42]. Then, the following HRV features are extracted using Vidaurre *et al.*'s *BioSig* toolbox [43] which provides methods for biomedical signal processing: difference between consecutive RR intervals (Root Mean Square of the Successive Differences–RMSSD), power spectral density (PSD) for Low Frequency (LF), PSD for High Frequency (HF), the ratio of LF to HF, and the total power. A total of 71 features are computed from the captured ECG signal and are concatenated into the final feature vector F_{ECG}.

3) Fusion of EEG and ECG-based features: The use of features based on multiple modalities has been shown to provide increased classification accuracy compared to approaches based on a single modality [1], [2], [11], [12]. In order to evaluate the performance of the combined EEG and ECG-based features, the two feature vectors F_{EEG} and F_{ECG} are fused as follows: First, the values of each feature vector are normalised in the range [0, 1] in order to compensate for the differences in numerical range. Then, the two normalised feature vectors F'_{EEG} and F'_{ECG} are concatenated in the final feature vector $F_{fused} = [\quad F'_{EEG} \quad F'_{ECG} \quad]$.

C. Baseline normalisation of features

Feature normalisation methods are commonly employed in machine learning approaches in order to make the data that form the feature vectors comparable, since their magnitude and range heavily depend on particular conditions, source signal characteristics and different subjects. The magnitude of the features extracted from the EEG and ECG signals varies significantly depending on the type of feature and its source (refer to Table V for the type of features). For example, when extracting features from EEG signals, the PSD for higher frequencies has a much smaller magnitude than the PSD for lower frequencies [32]. Physiological signals tend to have high variance between different subjects, as well as between the same features measured at different moments for each individual subject [44].

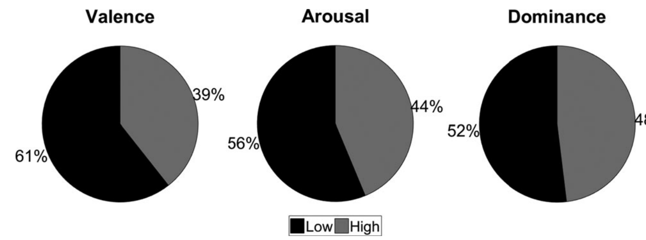

Fig. 5. Overall class distribution across all participants after conversion to a two-class rating score.

Furthermore, different types of features are measured in different units leading to variable numerical ranges across the features of a feature vector. Feature normalisation is employed in order to address these issues.

Baseline features are computed from the EEG and ECG recordings of the last $4\ s$ of the neutral film clip shown before each affect eliciting film clip. Then, each EEG and ECG-based feature is divided by the corresponding baseline feature. Due to the division of the extracted features by the baseline features, the derived normalised feature contains information about the relation of the initial feature to the background activity, i.e. activity that is present in the data but is not modulated by actual affective stimuli [32]. As a result, the feature normalisation method employed, attempts to remove or strongly attenuate the background activity in order to obtain only stimuli-related changes in the EEG and ECG recordings. It must be noted that in the case of fusion between EEG and ECG-based features, baseline normalisation is performed before the fusion procedure described in the respective Section.

V. Affect Recognition Results

In order to provide some baseline classification results for the proposed database, supervised classification experiments for affect recognition were conducted using the features extracted by the two modalities, as well as their fusion. Three different binary classification schemes were defined: the classification between low/high arousal (calm/excited), low/high valence (unpleasant/pleasant), and low/high dominance (without control/empowered). The subjective evaluations of each participant in the form of ratings using a scale from 1 to 5 were used as the ground truth, with the ratings being thresholded into two classes (low and high). Since a 5-point rating scale was used, the threshold was placed in the middle, leading to unbalanced classes for some participants and scales, as also reported by Koelstra *et al.* [12] in their work. The overall class distribution across the whole dataset for the transformed two-class rating scale is shown in Fig. 5. The dataset is balanced for dominance, with samples from the first class (low) constituting on average the 52% of the total samples, while samples from the second class (high) account for 48% of the total samples on average, with a standard deviation of 14.4%. The dataset is slightly unbalanced for arousal, with samples amounting for 56% and 44% on average for the first and second classes respectively, with a standard deviation of 14.2%. The dataset is more unbalanced for valence, where the first class amounts for 61% of the samples on average (39% for the second class), with a standard deviation of 10.6%.

To cope with the unbalanced class distribution for some participants, F1-scores are also reported alongside the classification accuracy in order to provide a more reliable measure of the classification success. Classification was performed independently for each participant in this study, using a Support Vector Machine (SVM) classifier with a Radial Basis Function (RBF) kernel. Matlab's SVM implementation was utilised for tuning, training and testing the SVM classifier. Furthermore, a 10-fold cross validation technique was used in order to validate the user independent classification performance. The eighteen samples for each participant were randomly divided into ten groups, leading to eight groups of two samples and two groups of one sample. At each step of the cross-validation, one group was used for testing and the other groups for training the classifier. This process was repeated until all groups had been used for testing. Furthermore, in order to compensate for the random selection of the samples at each group, the whole procedure was repeated ten times and the average result for all the iterations is reported. It must be noted that the 3-NN, 5-NN, 7-NN, LDA, and SVM with linear kernel classifiers were also evaluated using the same procedure, but failed to produce statistically significant results for all the examined cases (features and metrics).

Table VI presents the average accuracies and average F1 scores for both classes over all participants and for all rating scales. Classification accuracy for valence reached 62.49% using the EEG-based features, while the fusion of the EEG and ECG-based features provided the highest accuracy for arousal (62.32%). For dominance, the highest accuracy reached 61.84% for both the EEG-based and the fused features. In terms of F1 score, ECG-based features provided the best result for valence and arousal (0.5305 and 0.5798 respectively), while the fused features provided the best result for dominance (0.6171). Apart from the three proposed modalities (EEG, ECG, and their fusion), Table VI also reports the analytically computed results for using a random classifier with uniform distribution and for using a classifier that votes for each class according to the probability of the respective class to occur in the training data. It must be noted that the results of the random and class ratio classifiers are marginally overestimated since the actual class ratio should be determined from the training set used at each fold of the 10-fold cross validation. The random classifier results in an expected accuracy of 50% and an expected F1 score of approximately 0.49 for each participant, while classification according to the class ratio leads to an accuracy of approximately 54.4% and an F1 score equal to 0.50. The reported results were tested for statistical significance using both a one-way ANOVA and a Wilcoxon signed rank test, by comparing the accuracy and F1 scores over participants to the analytically computed expected results from voting according to the class ratio. Both significance tests resulted in p-values smaller than 0.002 for accuracy and F1 scores of all modalities examined and for all rating scales. Furthermore, the same procedure was followed for evaluating the statistical significance of the reported results compared to the expected results for random voting, leading to p-values smaller than 0.0001 for both accuracy and F1 score and for all rating scales.

Apart from the affect recognition results for the proposed database, Table VI also shows the baseline results reported for the DEAP [12], MAHNOB-HCI [1] and DECAF [2] databases, when using similar modalities for feature extraction (results for using eye gaze features or features from multimedia content analysis have been omitted, while results for dominance were not reported in the DEAP and MAHNOB-HCI studies). The baseline results for the proposed database are consistent with the baseline results reported for the other databases, with the exception of the F1 score for dominance which is elevated in the present study (0.6171 compared to 0.5300 for the DECAF database). Furthermore, for the DECAF database, the Magnetoencephalogram (MEG) modality was utilised instead of EEG, but results suggested that the affect encoding power of EEG and MEG is comparable, while MEG offers increased spatial resolution [2].

The comparable baseline results achieved for DREAMER and the other three databases provide a strong indication that the use of low-cost wireless devices is a viable alternative to expensive and non-portable medical equipment for capturing EEG and ECG signals for affect recognition applications. The fairest comparison between the devices used and medical grade devices would entail simultaneous recording with each modality for identical subjects and stimuli. Nevertheless, such study is probably impossible to implement in practice, as argued by Abadi *et al.* [2], thus the comparison of emotion recognition performance is based on the results observed on different populations (DREAMER, DEAP, MAHNOB-HCI, DECAF) using similar feature extraction methods when applicable.

VI. Discussion

As shown in Table VI, classification accuracy and F1 scores achieved for all modalities are close for each rating scale and within the statistical error. A Wilcoxon signed-rank test showed that the difference between the performance of the fusion strategy compared to the results of the EEG-based features and the ECG-based features was not significant for both accuracy and F1 score. This finding suggests that the fusion of EEG and ECG-based features is not beneficial for the affect recognition problem examined. Moreover, a correlation analysis of inter-participant results for EEG and ECG-based features using the Spearman correlation led to a Spearman's $\rho > 0.97$ for classification accuracy for all rating scales and a $\rho > 0.90$ for F1 scores, indicating a very strong correlation between the results. Furthermore, a Wilcoxon signed-rank test showed that the difference between the performance of the EEG and the ECG-based features was not significant. These findings suggest that the features computed from the two modalities provide similar descriptive power in relation to affect. Nevertheless, it is not clear whether this effect is a result of interference to the EEG signal from muscular and cardiac activity that is also captured in the ECG signal, or a result of lower descriptive power from the EEG due to the small number of electrodes of the wireless Emotiv EPOC device used, or a combination of these and other unpredicted factors.

After examining the mean classification accuracy achieved by each participant for each rating scale, some outliers across

TABLE VI
ACCURACY AND F1 SCORES OF AFFECT RECOGNITION FOR THE EXAMINED MODALITIES AND THE BASELINE FOR THE PROPOSED DATABASE, AND FOR OTHER WORKS PROPOSED IN THE LITERATURE USING OTHER DATABASES

	Accuracy			F1 Score		
Modality	Valence	Arousal	Dominance	Valence	Arousal	Dominance
EEG	**0.6249**	0.6217	**0.6184**	0.5184	0.5767	0.6166
ECG	0.6237	0.6237	0.6157	**0.5305**	**0.5798**	0.6145
Fusion (EEG & ECG)	0.6184	**0.6232**	**0.6184**	0.5213	0.5750	**0.6171**
Random	0.5000	0.5000	0.50.00	0.4885	0.4878	0.4895
Class ratio	0.5440	0.5467	0.5403	0.5000	0.5000	0.5000
DEAP EEG [12]	0.5760	0.6200	N/A	0.5630	0.5830	N/A
DEAP Peripheral [12]	0.6270	0.5700	N/A	0.6080	0.5330	N/A
MAHNOB-HCI EEG [1]	0.5700	0.5240	N/A	0.5600	0.4200	N/A
MAHNOB-HCI Peripheral [1]	0.4550	0.4620	N/A	0.3900	0.3800	N/A
DECAF MEG [2]	0.5900	0.6200	0.6200	0.5500	0.5800	0.5300
DECAF Peripheral [2]	0.6000	0.5500	0.5000	0.5900	0.5400	0.5000

Results in bold indicate the best results for the proposed database only. Peripheral signals include multiple modalities apart from ECG.

all modalities were detected. Outliers were considered the participants for which the classification accuracy was less than the expected mean accuracy for voting according to the class ratio for both the EEG and ECG-based features, as well as for their fused features. Two outliers were detected for valence, participant #1 and #19, for whom the mean classification accuracy reached 45.92% and 45.18% respectively. It is worth mentioning that both participants achieved very high classification accuracy for arousal compared to the mean accuracy achieved for all participants (62.29% overall vs 72.22% and 77.78% for participant #1 and #19 respectively), while participant #19 also achieved higher than the overall mean accuracy for dominance (61.75% overall vs 66.67%) and participant #1 was marginally not considered as an outlier for dominance (accuracy 55.56%). Three outliers were detected for arousal, participants #2, #15, and #18, for whom the mean classification accuracy reached 43.89%, 43.52% and 45.18% respectively. Participant #2 was also an outlier for dominance (accuracy 42.40%) but achieved very high classification accuracy for valence (72.22% vs 62.29% overall). Participant #15 achieved high accuracy for both valence and dominance (72.22% and 77.78% compared to 62.29% and 61.75% overall). Finally, participant #18 achieved higher than the overall mean accuracy for valence (62.29% overall vs 66.67%), while he was marginally not considered as an outlier for dominance (accuracy 55.56%). As a result, while outliers can be detected for individual rating scales, this behaviour does not extend to all scales and thus they cannot be considered as general outliers and be removed from the dataset. It is worth mentioning that all the aforementioned accuracies refer to the mean accuracy of the three feature sets used.

Concerning the usefulness of the proposed database, to the best of the authors knowledge, DREAMER is the first database which includes recordings from low-cost, off-the-shelf, portable, wireless devices for EEG and ECG signals. These devices have the potential to be used in non-professional everyday scenarios, providing the opportunity for affect recognition applications to reach a far wider user audience and for integration with multiple and diverse applications. As a result, this database can be of great value to researchers or industry members from multiple disciplines, including affective computing, multimedia, human-computer interaction, psychology, etc., as it provides the opportunity to study the various relations between affect and central and peripheral nervous system reactions. Furthermore, the proposed database allows the study of affect recognition methods on signals captured by low-cost and easy-to-use devices, increasing the interest for parties trying to introduce affect computing on a wide variety of applications.

VII. CONCLUSION

In this work, a database for the analysis of emotions elicited by audio-visual stimuli is presented. The database includes EEG and ECG recordings from 23 participants, where each participant rated his emotional response along the scales of valence, arousal, and dominance, after watching each of the 18 film clips selected to elicit specific emotions. The recordings of the EEG and ECG signals for this database were done using portable, wireless, low-cost, off-the-self devices that would allow for the integration of affective computing technology and algorithms into a wide range of applications. The quality of the participants ratings was evaluated through correlation analysis with the ratings of another study that utilised the same audio-visual stimuli, proving that there were significant correlations and low variability between the ratings. Moreover, participant-wise classification experiments for the scales of valence, arousal, and dominance were conducted in order to establish baseline results for the proposed database in terms of classification accuracy and F1 scores. The classification results using EEG and ECG-based features, as well as their fusion, were significantly higher than the results for random voting or voting according to the class ratio and they were also consistent with results in similar works that utilised non-portable medical grade equipment. These findings further support the argument that the use of low-cost off-the-shelf EEG and ECG devices for affect recognition applications is a viable alternative to expensive and non-portable medical equipment, a fact that can facilitate the integration of affect computing methods to everyday applications. The DREAMER database will be made publicly available after the publication of this work in order to give the opportunity to researchers to evaluate their algorithms on a database created from off-the-shelf wireless EEG and ECG devices and examine the possibility of applying them to general applications.

Acknowledgment

The authors would like to thank T. Cuntz and S. Palke for the data collection under their BSc (Hons) project.

References

[1] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A multimodal database for affect recognition and implicit tagging,” *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.
[2] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, “DECAF: MEG-based multimodal database for decoding affective physiological responses,” *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 209–222, Jul. 2015.
[3] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.
[4] P. Ekman *et al.*, “Universals and cultural differences in the judgments of facial expressions of emotion,” *J. Personality Soc. Psychol.*, vol. 53, no. 4, pp. 712–717, Oct. 1987.
[5] P. Ekman, “Strong evidence for universals in facial expressions: A reply to Russell’s mistaken critique,” *Psychol. Bull.*, vol. 115, no. 2, pp. 268–287, Mar. 1994.
[6] W. G. Parrott, *Emotions in Social Psychology: Essential Readings*. Hove, U.K.: Psychology Press, 2001.
[7] J. A. Russell and A. Mehrabian, “Evidence for a three-factor theory of emotions,” *J. Res. Personality*, vol. 11, no. 3, pp. 273–294, 1977.
[8] D. Watson, L. A. Clark, and A. Tellegen, “Development and validation of brief measures of positive and negative affect: The PANAS scales,” *J. Personality Soc. Psychol.*, vol. 54, no. 6, pp. 1063–1070, Jun. 1988.
[9] R. Plutchik, “The nature of emotions,” *Amer. Sci.*, vol. 89, no. 4, pp. 344–350, Jul./Aug. 2001.
[10] J. A. Russell, “A circumplex model of affect,” *J. Personality Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.
[11] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, “Analysis of EEG signals and facial expressions for continuous emotion detection,” *IEEE Trans. Affective Comput.*, vol. 7, no. 1, pp. 17–28, Jan. 2016.
[12] S. Koelstra *et al.*, “DEAP: A database for emotion analysis; using physiological signals,” *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.
[13] D. Bos, “EEG-based emotion recognition: The influence of visual and auditory stimuli,” Dept. Comput. Sci., Univ. Twente, 2007.
[14] B. R. Nhan and T. Chau, “Classifying affective states using thermal infrared imaging of the human face,” *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 979–987, Apr. 2010.
[15] H. Gunes and M. Piccardi, “A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior,” in *Proc. 18th Int. Conf. Pattern Recognit.*, 2006, vol. 1, pp. 1148–1153.
[16] G. Fanelli, J. Gall, H. Romsdorfer, T. Weise, and L. V. Gool, “A 3-D audio-visual corpus of affective communication,” *IEEE Trans. Multimedia*, vol. 12, no. 6, pp. 591–598, Oct. 2010.
[17] M. Grimm, K. Kroschel, and S. Narayanan, “The vera am mittag german audio-visual emotional speech database,” in *Proc. 2008 IEEE Int. Conf. Multimedia Expo.*, Jun. 2008, pp. 865–868.
[18] H. Gunes and B. Schuller, “Categorical and dimensional affect analysis in continuous input: Current trends and future directions,” *Image Vis. Comput.*, vol. 31, no. 2, pp. 120–136, 2013.
[19] M. A. Nicolaou, H. Gunes, and M. Pantic, “Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space,” *IEEE Trans. Affective Comput.*, vol. 2, no. 2, pp. 92–105, Apr. 2011.
[20] O. AlZoubi, S. K. D’Mello, and R. A. Calvo, “Detecting naturalistic expressions of nonbasic affect using physiological signals,” *IEEE Trans. Affective Comput.*, vol. 3, no. 3, pp. 298–310, Jul. 2012.
[21] C. A. Gabert-Quillen, E. E. Bartolini, B. T. Abravanel, and C. A. Sanislow, “Ratings for emotion film clips,” *Behav. Res. Methods*, vol. 47, no. 3, pp. 773–787, 2015.
[22] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, “Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers,” *Cogn. Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.
[23] N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur, “Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs,” *PeerJ*, vol. 1, p. e38, Feb. 2013. [Online]. Available: https://doi.org/10.7717/peerj.38
[24] H. Ekanayake, “P300 and Emotiv EPOC: Does Emotiv EPOC capture real EEG?” 2015. [Online]. Available: http://neurofeedback.visaduma.info/EmotivResearch.pdf.
[25] A. Burns *et al.*, “SHIMMER a wireless sensor platform for noninvasive biomedical research,” *IEEE Sensors J.*, vol. 10, no. 9, pp. 1527–1534, Sep. 2010.
[26] J. Wagner, J. Kim, and E. Andre, “From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification,” in *Proc. 2005 IEEE Int. Conf. Multimedia Expo.*, Jul. 2005, pp. 940–943.
[27] J. D. Morris, “Observations: Sam: The self-assessment manikin; an efficient cross-cultural measurement of emotional response,” *J. Advertising Res.*, vol. 35, no. 8, pp. 63–38, Jan. 1995.
[28] MathWorks, Inc., “MATLAB,” Natick, MA, USA, 2016. [Online]. Available: http://www.mathworks.com.
[29] M. M. Bradley and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential,” *J. Behav. Therapy Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.
[30] A. Delorme and S. Makeig, “EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 2004.
[31] C. A. Kothe, “The Artifact Subspace Reconstruction method,” 2013. [Online]. Available: http://sccn.ucsd.edu/eeglab/plugins/ASR.pdf, Accessed on: Jul. 14, 2016.
[32] M. X. Cohen, *”Analyzing Neural Time Series Data: Theory and Practice*. London, U.K.: MIT Press, 2014.
[33] R. J. Davidson, “Affective neuroscience and psychophysiology: Toward a synthesis,” *Psychophysiology*, vol. 40, no. 5, pp. 655–665, Sep. 2003.
[34] H. W. Cole and W. J. Ray, “EEG correlates of emotional tasks related to attentional demands,” *Int. J. Psychophysiol.*, vol. 3, no. 1, pp. 33–41, Jul. 1985.
[35] R. J. Barry, A. R. Clarke, S. J. Johnstone, C. A. Magee, and J. A. Rushby, “EEG differences between eyes-closed and eyes-open resting conditions,” *Clin. Neurophysiol.*, vol. 118, no. 12, pp. 2765–2773, Dec. 2007.
[36] R. J. Barry, A. R. Clarke, S. J. Johnstone, and C. R. Brown, “EEG differences in children between eyes-closed and eyes-open resting conditions,” *Clin. Neurophysiol.*, vol. 120, no. 10, pp. 1806–1811, Oct. 2009.
[37] M. Soleymani and M. Pantic, “Multimedia implicit tagging using EEG signals,” in *Proc. 2013 IEEE Int. Conf. Multimedia Expo.*, Jul. 2013, pp. 1–6.
[38] P. Rainville, A. Bechara, N. Naqvi, and A. R. Damasio, “Basic emotions are associated with distinct patterns of cardiorespiratory activity,” *Int. J. Psychophysiol.*, vol. 61, no. 1, pp. 5–18, Jul. 2006.
[39] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, “Looking at pictures: Affective, facial, visceral, and behavioral reactions,” *Psychophysiology*, vol. 30, no. 3, pp. 261–273, May 1993.
[40] R. McCraty, M. Atkinson, W. A. Tiller, G. Rein, and A. D. Watkins, “The effects of emotions on short-term power spectrum analysis of heart rate variability,” *Amer. J. Cardiol.*, vol. 76, no. 14, pp. 1089–1093, Nov. 1995.
[41] J. Pan and W. J. Tompkins, “A real-time QRS detection algorithm,” *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–236, Mar. 1985.
[42] J. Wagner, “Augsburg biosignal toolbox (AuBT),” Univ.Augsburg, Augsburg, Germany, 2005.
[43] C. Vidaurre, T. H. Sander, and A. Schlogl, “BioSig: The free and open source software library for biomedical signal processing,” *Comput. Intell. Neurosci.*, vol. 2011, 2011, Art. no. 935364.
[44] D. Novak, M. Mihelj, and M. Munih, “A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing,” *Interact. Comput.*, vol. 24, no. 3, pp. 154–172, May 2012.