# AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups

Juan Abdon Miranda-Correa, *Student Member, IEEE,* Mojtaba Khomami Abadi, *Student Member, IEEE,* Nicu Sebe, *Senior Member, IEEE,* and Ioannis Patras, *Senior Member, IEEE*

**Abstract**—We present AMIGOS– A dataset for Multimodal research of affect, personality traits and mood on Individuals and GrOupS. Different to other databases, we elicited affect using both short and long videos in two social contexts, one with individual viewers and one with groups of viewers. The database allows the multimodal study of the affective responses, by means of neuro-physiological signals of individuals in relation to their personality and mood, and with respect to the social context and videos' duration. The data is collected in two experimental settings. In the first one, 40 participants watched 16 short emotional videos. In the second one, the participants watched 4 long videos, some of them alone and the rest in groups. The participants' signals, namely, Electroencephalogram (EEG), Electrocardiogram (ECG) and Galvanic Skin Response (GSR), were recorded using wearable sensors. Participants' frontal HD video and both RGB and depth full body videos were also recorded. Participants emotions have been annotated with both self-assessment of affective levels (valence, arousal, control, familiarity, liking and basic emotions) felt during the videos as well as external-assessment of levels of valence and arousal. We present a detailed correlation analysis of the different dimensions as well as baseline methods and results for single-trial classification of valence and arousal, personality traits, mood and social context. The database is made publicly available.

**Index Terms**—Emotion Classification, EEG, Physiological signals, Signal processing, Personality traits, Mood, Affect Schedules, Pattern classification, Affective Computing.

## 1 Introduction

Affective computing aims for the detection, modeling and synthesis of human emotional cues in Human-Computer Interaction [1]. In this field, an increasing interest has arisen for considering the user's affective responses when making computational decisions. For instance, Chanel et al [2] modified the difficulty of a video game according to user's emotional state to maintain engagement. In a hypothetical scenario, a movie time-line could be adapted to elicit specific affective states, based on factors such as viewer's predicted emotions, personality and mood. Hence, in these scenarios, it is very important to reliably predict such factors.

Advances on affective states prediction have been boosted by the availability of annotated affective databases, which act as benchmark for the developing of methodologies. These databases have used stimuli, such as music videos [1], short videos [3], [4], and diverse emotion elicitation methods [5]. They include information from different modalities (e.g. EEG, facial expression).

Available multimodal affective databases have studied affective responses of participants in individual [1], [6], or pairs of people/limited agent settings [7]. However, in real life, affective experiences are often performed in social contexts (e.g. movies and games are commonly engaged by groups of people together). In such contexts, the individual experiences do not depend only on the user and the content, but also on the implicit and explicit interactions that can occur between the personalities, reactions, moods and emotions of other group members. Additionally, different aspects of affect and personality could be inhibited or amplified depending on the social context of a person. Therefore, current databases have ignored an important dimension for the study of affect. There are databases that have been used for studying emotion in groups [8]–[10]. For instance Huang et al [10] showed that it is possible predict the group emotion. All these databases are limited to static images.

Databases for personality research have considered information related to linguistics in written text [11], social networks activity [12], and behavior in group activities [13]. However they have largely ignored the study of both, affect and personality, through the use of physiological signals, which have shown to carry valuable information for personality recognition [14], [15].

Therefore, there is a need of multimodal databases for the study of people's emotions, personality and mood, with subjects in both alone and group settings. The multimodal framework would benefit from the inclusion of neurological and peripheral physiological signals.

Our contribution to the field is A dataset for Multimodal research of affect, personality traits and mood on Individuals and GrOupS (AMIGOS) by means of neurophysiological signals. The dataset consists of multimodal recordings of participants and their responses to emotional fragments of movies. In our dataset: (i) The participants took part in two experiments where they watched one of two sets of stimuli, one of short videos and one of long videos,

*Juan Abdon Miranda-Correa and Ioannis Patras are with the School of Computer Science and Electronic Engineering, Queen Mary University of London, UK. E-mail: {j.a.mirandacorrea,i.patras}@qmul.ac.uk.*
*Mojtaba Khomami Abadi and Nicu Sebe are with the Department of Information Engineering and Computer Science, University of Trento, Italy. E-mail: {khomamiabadi,sebe}@disi.unitn.it.*

while their implicit responses, namely, Electroencephalogram (EEG), Electrocardiogram (ECG), Galvanic Skin Response (GSR), frontal HD video, and both RGB and depth full body videos were recorded. Recordings were precisely synchronized to allow the study of affect, personality and mood from the different modalities simultaneously. (ii) In the first experiment, all participants watched the set of short videos in individual setting. In the second experiment, some of the participants took part in individual setting and some of them in group settings. Then they watched the set of long videos. (iii) The participants have been profiled according to their personality through the Big-Five personality traits model, and according to their mood through the Positive Affect and Negative Affect Schedules (PANAS). (iv) Affective annotation was obtained with both internal and external methods. Internal annotation consisted of participants's self-assessment of affect at the experiment's beginning and immediately after each video. As external annotation, the recordings of both sets of videos were offline annotated by 3 annotators on both valence and arousal scales, using a method that allows the direct comparison of the affective responses from both experiments. (v) Physiological signals were recorded using commercial wearable sensors that allow more freedom for the participants than conventional laboratory equipment. The database is available to the academic community[1].

In this work, we present a comparison between the internal and external annotations of valence and arousal. We then perform correlation analysis between the affective responses elicited by short and long videos with respect to social context (whether a participant was alone or in a group) and between the participants' personality traits, PANAS and social context. We also present baseline methodologies and results for single-trial prediction of valence and arousal, and for prediction of personality traits, PANAS and social context, using neuro-physiological signals (EEG, ECG and GSR) as single modalities and fusion of them.

Our main findings are as follows: (i) We show that there is significant correlation between internal and external annotations indicating that external annotation is a good predictor of the affective state of participants. (ii) We show, by correlation analysis of external annotations, that in the eyes of annotators, participants seem to have low arousal in low valence moments and high arousal for high valence moments. (iii) We found significant differences in the distribution of valence and arousal, externally annotated, between individual settings compared to group settings for long videos. It was different for the short videos experiment where the distribution of arousal and valence for the 2 sets of participants are not statistically different ($p > 0.05$). This result was expected since all the participants watched the short videos within the same social context (alone). (iv) We found significant negative correlations between the scores of negative affect (NA) and the ones of extraversion, agreeableness, emotional stability and openness, and significant positive correlations between the scores of agreeableness and both extraversion and positive affect (PA), between consciousness and emotional stability, and between PA and arousal. Finally, (v) our method for personality traits, mood and social context prediction based on neuro-physiological signals outperforms a previous study [14] in prediction of extroversion, emotional stability, PA and NA usign EEG and in prediction of conscientiousness, openness and conscientiousness using physiological signals (ECG and GSR).

In section 2, we present works related to affect, personality and mood modeling and assessment, and a survey of main multimodal databases for affect and personality research. Section 3 presents the experimental scenarios, stimuli selection, modalities and equipment used to record the implicit responses. Then, an overview of the experimental setup and the methods employed for assessment of affect, personality traits and mood (PANAS) are described. In Section 4, the data obtained from the different experiments is analyzed. Section 5 presents our method for valence and arousal recognition as well as our approach for personality traits, PANAS and social context recognition using neuro-physiological signals. The results are then presented and discussed. Finally, we conclude in section 6.

## 2 Related Works

In this section, we make a review of the works related with modeling and assessment of affect, personality and mood. Next, we review of important databases that study them.

### 2.1 Affect, Personality and Mood

Plutchnik [16] defined emotion as a complex chain of loosely connected events that begins with a stimulus and includes feelings, psychological changes, impulses to action and specific, goal-directed behavior. Common approaches to model affect are categorical and dimensional. The former claims that there exists a small number of emotions that are basic and recognized universally; The most common of these models is the Six Basic Emotions model [17], that categorizes emotions into fear, anger, disgust, sadness, happiness and surprise. The dimensional approach considers that affective states are inter-related in a systematic way (*e.g.* the Plutchik's emotion wheel [16]). Russell [18] introduced the Circumplex Model of Affect, where affective states are represented in a two dimensional space with arousal (the degree an emotion feels active) and valence (the degree an emotion feels pleasant) as the main dimensions.

Affective experiences are also modulated by people's internal factors, such as mood and personality [19]. Personality refers to stable individual characteristics, that explain and predict behavior [20]. The Big-Five factor model [21] describes personality in terms of five traits (dimensions) namely Extraversion (sociable vs reserved), Agreeableness (compassionate vs dispassionate and suspicious), Conscientiousness (dutiful vs easy-going), Emotional stability (nervous vs confident) and Openness to experience (curious vs cautious). The common method to measure these dimensions is the use of questionnaires such as the Neuroticism, Extraversion and Openess Five Factor Inventory (NEO-FFI) [22] and the Big-Five Marker scale (BFMS) [21].

Mood refers to baseline levels of affect that define peoples experiences. It is commonly modeled using the two dimensions called Positive Affect (PA) and Negative Affect (NA) scales [23]. PA reflects the extent to which a person

1. http://www.eecs.qmul.ac.uk/mmv/datasets/amigos/

feels enthusiastic, active and alert. In contrast, NA is a general dimension of subjective distress and unpleasant engagement.In order to measure PA and NA, Watson et al [24] developed the Positive and Negative Affect Schedules (PANAS) that consist of two 10-item mood scales; These schedules have shown to be internally consistent, uncorrelated and stable over a 2-month time period.

### 2.2 Databases for Affective Computing

Databases for the study of affective computing have been developed to allow researchers to compare methods. Here, we review databases based on visual and neurophysiological signals modalities.

Databases for the study of affect recognition based on visual modality have focused mainly on the analysis of facial expressions. One of the main examples is the Sustained Emotionally Colored Machine-human Interaction using Nonverbal Expression (SEMAINE) database [7]. It consists of high-quality, multimodal recordings of 150 participants in emotionally colored conversations. It is annotated for valence, arousal and action units (AUs). Another example is the Affectiva-MIT Facial Expression Dataset (AM-FED) [25]. It is a labeled dataset of spontaneous facial responses recorded in natural settings on the Internet. The dataset consists of 242 facial videos, labels of the presence of 10 symmetrical and 4 asymmetrical AUs, 2 head movements, smile, general expressiveness, feature tracker fails, gender, location of 22 automatically detected landmark points and self-report responses of familiarity, liking and desire to watch again. The Denver Intensity of Spontaneous Facial Action (DISFA) database [26] consists of labeled stereo video recordings of 27 adults while watching a video clip. Labels consist of presence, absence and intensity of 12 facial AUs. Dhall et al [8] presented HAPPEI, a database containing 4886 images of groups collected in the wild and annotated for happiness intensity.

Databases based on physiological signals include the MAHNOB-HCI [6]. It is a multimodal database of synchronized recordings of face video, audio signals, eye gaze data and physiological signals (ECG, GSR, respiration amplitude (RA), skin temperature (ST) and EEG) of 27 participants while watching first, 20 videos, and second, short videos and images with relevant/non-relevant tags. It includes self-reports of arousal, valence, dominance, predictability scales, emotional keywords and agreement or disagreement with the tags. Koelstra et al present the DEAP database [1], that includes EEG and peripheral physiological signals (GSR, RA, ST, ECG, blood volume, Zygomaticus and Trapezius muscles Electromyogram and Electrooculogram) recordings. It includes video and signals' recordings of 32 participants while watching 40 music video clips. It includes self-assessment of arousal, valence, liking, dominance and familiarity. A similar database that uses Magnetoencephalogram (MEG) is the DECAF database, which includes recordings of 30 participants. More recently, Zhang et al [5] collected the Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis. It includes 140 participants exposed to 10 different emotion elicitation methods for surprise, disgust, fear, etc. Recorded signals are 3D and 2D videos, thermal sensing, electrical conductivity of the skin, respiration, blood pressure and hearth rate.

One database for personality research using video modality is the Mission Survival II corpus [13]. It is a multimodal annotated collection of video and audio recordings (4 cameras, 17 microphones) of four meetings, of 4 participants. Participants were profiled in terms of the Ten Item Personality Inventory [27] to account for their personality states (moments where participants act more or less extravert, creative, ect). Affect is not considered in this dataset. A recent multi-modal database for implicit personality and affect recognition is the ASCERTAIN [28]. It includes recordings of the EEG, ECG, GSR and facial video of 58 users, while viewing short movie clips. This database only includes participants in individual configuration and does not share data about mood of participants.

To the best of our knowledge there are not databases for personality research based on neurological or physiological signals and that studies participants in both individual and group settings. In Table 1, we summarize the characteristics of the reviewed databases and compare them to ours.

## 3 Experimental Setup

In this section, we describe experimental scenarios. Then, the process for selection of stimuli is explained, and modalities and equipment used are presented. Then, the experimental protocol is described in detail. Finally, the procedures for internal and external annotation of affect and for personality and mood assessment are introduced.

### 3.1 Experimental scenarios

The main objective of this work is to study the personality, mood and affective response of people engaging with multimedia content in two social contexts, (i) when they are alone (individual setting), and (ii) when they are part of an audience (group setting). At the same time, we study people's affective response to two types of eliciting content. First, short emotional videos (duration$<$250s) selected to elicit specific affective states in the participants. Second, long videos (duration$>$14min), that could elicit various affective states over their duration in which story and narrative could amplify affective responses. Therefore, we designed two experiments, in the first one (Short videos experiment), all participants watched short videos in individual setting. In the second experiment (Long videos experiment), the same participants watched long videos, some of them did it in individual setting, while the others did it in group setting.

### 3.2 Stimuli selection

Emotion elicitation depends greatly on a careful selection of the stimuli, which needs to be suitable for the objective of the study and allow for consistent results among trials [1]. In this work, we selected two sets of videos for emotion elicitation. The first one consists of short emotional videos and the second one of long videos. For the first set, 72 volunteers annotated, on the valence and arousal dimensions, the set of 36 videos used in [3]. We then classified each of the videos into one of four quadrants of the valence-arousal (VA) space, namely HVHA, HVLA, LVHA and LVLA (H, L, A and V stand for high, low, arousal and valence respectively). From each quadrant, we selected the three videos laying further

TABLE 1
Summary of characteristics of databases for affect and personality. Last row is our database.

| Database | No. Part. | Individual vs. Group | Purpose | Modalities | Annotations |
|---|---|---|---|---|---|
| **SEMAINE** [7] | 150 | Individual | Emotion recognition based on facial expressions | Audio and Visual | Valence, arousal and FACS. |
| **AM-FED** [25] | 242 | Individual | Spontaneous facial expression recognition "In-the-Wild" | Visual | 14 AUs, 2 head movements, smile, expressiveness and 22 landmark points. Self-assessment of familiarity, liking and desire to watch again. |
| **DISFA** [26] | 27 | Individual | Spontaneous facial action recognition | Visual | 12 AUs. |
| **HAPPEI** [8] | - | Group (4886 images) | Group happiness intensity research | Visual (Facial Expressions) | Group level mood intensity ('neutral' to 'thrilled'), face level happiness intensity, occlusion intensity and pose. |
| **MAHNOB-HCI** [6] | 27 | Individual | Emotion recognition and implicit tagging | Visual, Audio, Eye Gaze, ECG, GSR, Respiration Amplitude, Skin temperature, EEG. | Self-assessment of valence, dominance, predictability and emotional keywords. Agreement/disagreement with tags. |
| **DEAP** [1] | 32 | Individual | Implicit affective tagging from EEG and peripheral physiological signals | EEG, GSR, Respiration Amplitude, Skin Temperature, Blood Volume, Electromyogram and Electrooculogram. Visual for 22 participants. | Self-assessment of arousal, valence, liking, dominance and familiarity. |
| **DECAF** [3] | 30 | Individual | Affect recognition | MEG, Near-infra-red facial video, horizontal Electrooculogram, ECG and trapezius-Electromyogram. | Self-assessment of valence, arousal and dominance. Continuous annotation of valence and arousal of the stimuli. |
| **Zhang et al corpus** [5] | 140 | Individual | Emotional behaviour research | 3D dynamic imaging, Visual, Thermal sensing, EDA, Respiration, Blood Pressure and Hearth Rate. | Occurrence and intensity of AUs. Features from 3D, 2D and Infra-red sensors. |
| **Mission Survival II** [13] | 16 | 4 people group | Personality states research | Audio and Visual | Personality states by the Ten Item Personality Inventory. |
| **ASCERTAIN** [28] | 58 | Individual | Personality and Affect | EEG, ECG, GSR and Visual | Big-Five personality traits, self-assessment of valence and arousal. |
| **AMIGOS** | 40 | Individual & 4 people group | Affect, personality, mood and social context recognition | Audio, Visual, Depth, EEG, GSR and ECG | Big-Five personality traits and PANAS. Self-assessment of valence, arousal, dominance, liking, familiarity and basic emotions. External annotation of valence and arousal. |

TABLE 2
The short videos listed with their sources (Video IDs are stated in parentheses). In the category column, H, L, A and V stand for high, low, arousal and valence respectively.

| Category | Excerpt's source |
|---|---|
| **HAHV** | Airplane (4), When Harry Met Sally (5), Hot Shots (9), Love Actually (80) |
| **LAHV** | August Rush (10), Love Actually (13), House of Flying Daggers (18), Mr Beans' Holiday (58) |
| **LALV** | Exorcist (19), My girl (20), My Bodyguard (23), The Thin Red Line (138) |
| **HALV** | Silent Hill (30), Prestige (31), Pink Flamingos (34), Black Swan (36) |

TABLE 3
Selected Long Videos with Their ID, Source (Movie title. Director. Producer company. Released Year.) and Excerpt Duration. Note: Exact time-stamps of the excerpts are available at the dataset website.

| ID | Source | Duration |
|---|---|---|
| N1 | The Descent. Dir. Neil Marshall. Lionsgate. 2005. | 23:35.0 |
| P1 | Back to School Mr. Bean. Dir. John Birkin. Tiger Aspect Productions. 1994. | 18:43.0 |
| B1 | The Dark Knight. Dir. Christopher Nolan. Warner Bross. 2008. | 23:30.0 |
| U1 | Up. Dirs. Pete Docter and Bob Peterson. Walt Disney Pictures and Pixar Animation Studios. 2009. | 14:06.0 |

to the origin of the scale, totaling 12 videos. Additionally, from the videos used in [6], we selected four videos, each corresponding to one of the four quadrants. The total number of selected short videos is 16, 4 for each VA quadrant. We preserved the IDs used in the original datasets. Selected short videos (51-150s long, $\mu = 86.7$, $\sigma = 27.8$) with their corresponding category and their IDs are listed in Table 2.

For the second set of videos, we initially selected 8 video extracts from movies based on their score in the IMDb Top Rated Movies list[2]. We selected movies that could allow us to extract a long segment ($\approx$ 20min) which could be self-contained, did not require previous knowledge from the participants to be understood and with strongly affective multimedia content (good combination of music and colors [29]). Four researchers classified them as belonging to one or more quadrants of the VA space. Finally, 4 videos were selected favoring the extracts that could evoke emotions in different quadrants of the VA space, and making sure all the quadrants were covered. The selected long videos (14.1-23.58min, $\mu = 20.0$, $\sigma = 4.5$) with their corresponding video ID, source and duration are listed in Table 3.

2. http://www.imdb.com/chart/top

### 3.3 Neuro-Physiological Signals and Instruments

We recorded three main neuro-physiological signals namely EEG, ECG and GSR, which have shown good performance in affect estimation studies [30]–[32]. We opted for these modalities because they allow us to use only wearable sensors, which would let the users feel as comfortable as possible. Below we give an introduction of each of them.

**EEG**: Electroencephalogram is a recording of the electrical activity along the scalp. It measures voltage fluctuations resulting from ionic current flows within the brain [33]. EEG signals carry valuable information about the person's affective state [34]–[36].

**GSR**: Galvanic skin response, also known as electrodermal activity (EDA), measures the electrical conductance of the skin [37], [38]. Skin conductivity varies with changes in skin moisture level (sweating) which can reveal changes in autonomous nervous system (ANS) related to arousal [32], [39], [40], revealing emotions such as stress or surprise [39].

**ECG**: Electrocardiogram is a recording of the electrical activity of the heart generated by the polarization and depolarization of cardiac tissue. It is detected by electrodes attached to the skin surface. ECG can reveal changes of the ANS related to affective experiences and stress [30], [41].

In previous databases, neuro-physiological signals have been recorded using laboratory equipment (e.g. Biosemi ActiveTwo) which is expensive and limits the mobility of the participants. In this database we use wearable sensors that allow more freedom given that they use wireless technology. EEG was recorded using the Emotiv EPOC Neuroheadset[3] (14 channel, 128 Hz, 14 bit resolution). EEG channels according to the 10-20 [42] system are: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. This headset has been used previously for affect recognition [43], [44]. ECG was recorded using the Shimmer 2R[4] platform extended with an ECG module board (256 Hz, 12 bit resolution), which uses three electrodes, two of them are placed at the right and left arm crooks and the third one at the internal face of the left ankle as reference. This set-up allows precise identification of heart beats as well as the full ECG QRS complex. GSR signal was recorded using the Shimmer 2R platform extended with a GSR module board (128 Hz, 12 bit resolution), with two electrodes placed at the middle phalanges of the left hand's middle and index fingers.

### 3.4 Video Recordings

Video modality is widely used for assessing peoples affective states [25], [45], [46]. Frontal face video was recorded in HD quality using a JVC GY-HM150E camera, positioned just below the screen. Additionally, both RGB and depth full body videos were recorded using a Microsoft's Kinect V1[5] placed at the top of the screen. A participant during the short videos experiment and a group of participants during the long videos experiment can be observed in Fig. 1.

### 3.5 Synchronization and Stimuli Display Platform

One PC (Intel Core i7, 3.4 GHz) was used to (i) present the stimuli, (ii) get and synchronize signals, and (iii) obtain the self-assessment of participants. Shimmer sensors were paired to the PC using bluetooth standard, while the Emotiv headset was paired using a proprietary wireless standard. Videos were presented in a 40-inch screen (1280×1024), each of them was displayed preserving the original aspect ratio and covering the highest screen-area possible. The remaining area was filled with black background. Subjects were seated approximately 2 meter from the screen. Stereo speakers were used and the sound volume was set at a relatively loud level, however it was adjusted when necessary.

### 3.6 Short Videos Experiment Protocol

Recordings were performed in a laboratory environment with controlled illumination. 40 healthy participants (13 female), aged between 21 and 40 (mean age 28.3), took part in the experiment. Prior to the recording session, the participants read and signed a consent form. Then they read a sheet with instructions about the experiment.When the instructions were clear, the participants were led into the experiment room. After that, the experimenter explained the affective scales used in the experiment and how to fill in the self-assessment form (See 3.8.1). Next, the sensors were placed and the signals quality was assessed. Finally, the experimenter left the room and the session began.

The participants performed an initial self-assessment for arousal, valence and dominance, and selected basic emotions (Neutral, Happiness, Sadness, Surprise, Fear, Anger and Disgust) they felt before any stimulus have been shown. Next, 16 videos were presented in a random order in 16 trials, each consisting of: (1) A 5 second baseline recording showing a fixation cross. (2) The display of a small video. (3) Self-assessment of arousal, valence, dominance, liking and familiarity as well as selection of basic emotions (See 3.8.1). After the 16 trials, the recording session ended.

3. http://www.emotiv.com/
4. http://www.shimmersensing.com/
5. http://developer.microsoft.com/en/windows/kinect/hardware

TABLE 4
Participant IDs for Individual and Group Settings of the long videos experiment. In the group setting, the IDs order represent the order in which participants were seated, from a front view, from left to right.

| | Part. ID | | Part. ID |
|---|---|---|---|
| Group 1 | 7, 1, 2, 16 | Group 5 | 15, 11, 12, 10 |
| Group 2 | 6, 32, 4, 3 | Individual Participants | 9, 13, 19, 20, 23, 25, 26, 30, 21, 33, 34, 35, 36, 37, 38, 39, 40 |
| Group 3 | 29, 5, 27, 21 | | |
| Group 4 | 18, 14, 17, 22 | | |

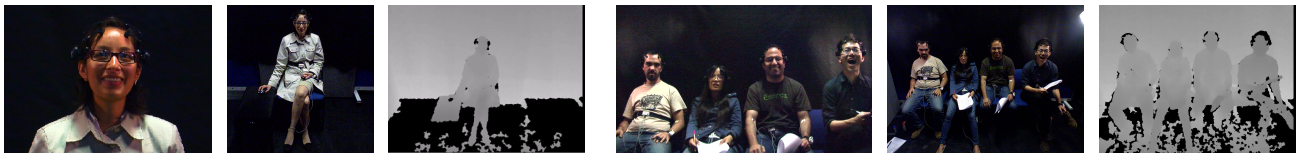
### 3.7 Long Videos Experiment Protocol

37 participants that took part in the short videos experiment, performed the long videos experiment in either individual or group settings (participants 8, 24 and 28 were not available). In the individual setting, 17 participants performed the experiment alone. In the group setting, 20 participants performed the experiment together with 3 other participants (5 groups of 4 people). In order to maximize interactions, groups were formed to include people that knew each other, being either friends, colleagues, or people with similar cultural background [47]. The IDs of participants that were in the individual or group setting are listed in Table 4.

During the recording sessions, the participant(s) was(were) led to the recording room. While the different sensors were set up, experimenters explained the differences of the protocol compared to the short videos experiment. Every participant was given a set of self-assessment paper forms (See 3.8.1) and a pen for self-assessment. Experimenters avoided to mention whether the participants could talk during the experiment, for the interactions to be spontaneous. Once the sensors had been tested, the experimenters left the room and the recording session started.

The experiment consisted of the display of 4 long videos in random order. Videos were shown in two recording subsessions, each consisting of: (1) initial self-assessment (45s) of arousal, valence, dominance and selection of basic emotions. (2) the display, in two trials, of two long videos, each followed by (3) self-assessment (45s) of arousal, valence, dominance, liking and familiarity, and selection of basic emotions (See 3.8.1). After the first sub-session followed a break of 15 minutes where participants were offered refreshments. After, sensors' signals were checked and the second recording sub-session started, after which the session ended.

After the long videos experiment, participants were asked to fill in as soon as possible, on-line forms with Personality Traits [48] and PANAS [24] questionnaires (See 3.9). Participants took 2 days on average to fill in the forms. Once they filled in all required forms, they were given mugs and university gadgets in return for their participation.



(a) (b) (c) (d) (e) (f)

Fig. 1. Participant in experiment conditions during the short videos experiment recorded in (a) Frontal HD video, (b) full body RGB video via Kinect, (c) full body depth video via Kinect; and group of 4 participants during the long videos experiment recorded in (d) frontal HD video, (e) full body RGB video via Kinect and (f) full body depth video via Kinect.

### 3.8 Affective Annotation

Internal annotation (self-assessment) is the process where a subject directly assess its affective state [49]. It has the advantage of being easy and a direct way to assess affective states. At the same time, it is an intrusive process, subjects could be unreliable at reporting their emotions or they could hide their real emotions [50]. External annotation (implicit assessment) is a process that intends to assess a person's affective state by external (indirect) means such as analyzing the person's behavior and/or its physiological responses [6]. We have performed both internal and external annotations to assess the participants' affective state.

#### *3.8.1 Participant's Affect Self-assessment*

At the beginning of the recording session of the short videos experiment, and of each of the two recording sub-sessions of the long videos experiment, participants performed self-assessment of their levels of arousal, valence and dominance, and selected basic emotions that described their emotions at the start of each session. Then, at the end of each trial, participants performed self-assessment of the same dimensions, and of the liking and familiarity that described what they felt during each video.

The self-assessment form used for the short videos experiment can be seen in Fig. 2. Self-assessment manikins (SAM) [51] were used to visualize the scales of valence, arousal and dominance. For the liking scale, thumbs down/thumbs up symbols were used. The fifth scale asks the participants to rate their familiarity with the video. Arousal scale ranges from "very calm" (1) to "very excited" (9). Valence from "very negative" (1) to "very positive" (9). Dominance from "overwhelmed with emotions" (1) to "in full control of emotions" (9). The fourth scale ranges from disliking (1) to liking (9) the video. The familiarity scale ranges from "Never seen it before" (1) to "Know the video very well" (9). Participants moved a continuous slider, placed at the bottom of each scale, to specify their self-assessment level. They could move the slider anywhere directly below or in-between of the manikins. Finally, participants were asked to select at least one of the basic emotions (Neutral, Disgust, Happiness, Surprise, Anger, Fear and Sadness [17]), or as many as they felt during the video (a participant can consider a video to be both happy and sad).

In the long videos experiment, having a digital form for every participant of the groups was not practical, therefore we opted to use a paper version of the form in Fig. 2 in both individual and group setting recordings, in order to keep consistent the self-assessment between settings.

In total, for the short videos experiment 17 annotations were obtained from each participant (1 at the beginning of the experiment and 1 after each of the 16 short videos), and 6 annotations in the case of the long videos experiment (1 at beginning of the first sub-session, 1 after each of the two long videos of the first sub-session, 1 at the beginning of the second recording sub-session just after the 15 minute break and 1 after each of the two long videos of the second sub-session). It is important to note that this annotation gives information related only to the participants' initial and final affective states, not for specific instants during the videos.

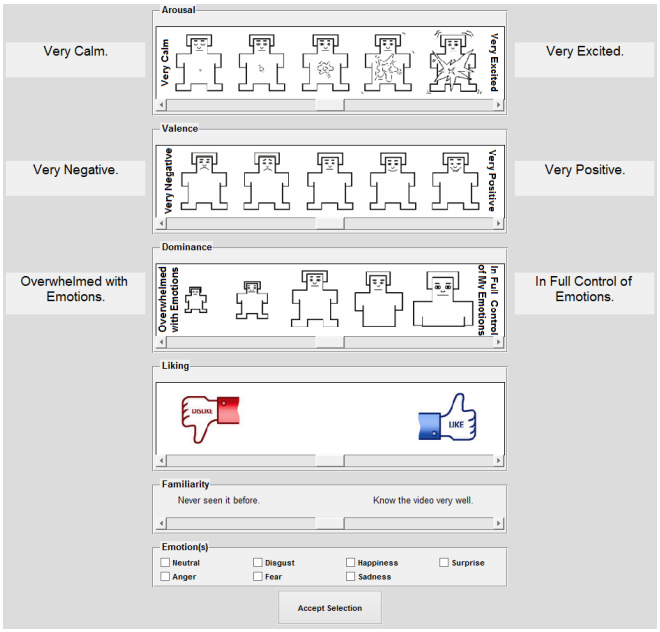



Fig. 2. Self-Assessment Form for Assessment of Arousal, Valence, Dominance, Liking, Familiarity and Basic Emotions.

#### *3.8.2 External Affect Annotation*

In order to study the temporal evolution of affect, the frontal videos of each participant recorded during the display of the stimuli of both experiments were off-line annotated on the valance and arousal dimensions as follows.

First, the videos of a given participant recorded during the display of each of the 20 stimuli videos (16 short and 4 long), were manually cropped in order to show only a squared region around the face, covering from the top of the head to the start of the shoulders. Then, each of the videos were split into 20 second clips. For this, the first 20 seconds of each video, including 5 seconds prior to the presentation of the stimuli, were extracted as first clip, then, starting from the 5s of the video (instant in which the stimuli started), $n = \lfloor (D)/(20s) \rfloor$ non overlapping segments of 20s were

extracted, with $D$ being the duration of the stimuli video in seconds. Finally, the last 20 seconds of the video were extracted as final clip. For every participant, {6, 7, 5, 6, 4, 5, 8, 5, 7, 5, 9, 5, 5, 4, 6, 7, 72, 58, 72 and 44} clips were obtained from videos {4, 5, 9, 10, 13, 18, 19, 20, 23, 30, 31, 34, 36, 58, 80, 138, N1, P1, B1 and U1}, totaling 340 clips per participant, 94 from the short videos and 246 from the long videos.

Three annotators rated on the valence and arousal scales the clips of all the participants (340 clips $\times$ 37 participants $=$ 12580 clips). Both scales were continuous and ranged from $-1$ (low valence/arousal) to 1 (high valence/arousal). The 340 clips of a given participant, were annotated in the same random order by each annotator, however, the order of the clips was different for each participant. Since samples of both experiments were randomly shown to the annotators, labels of the two experiments are directly comparable. The pipeline of the annotation consisted of the display of a randomly selected clip followed by the annotation performed by the annotator, first, of valence and then of arousal. This process was repeated until all clips were annotated.

### 3.9 Personality and Mood Assessment

The Big-Five personality traits were measured with an on-line form of the big-five marker scale questionnaire [21], in which, for each personality trait, using the basic question "I see myself as a person:", ten descriptive adjectives are rated with a 7-point-likert-scale [52] and a mean is calculated.

Mood was assessed on the positive affect (PA) and negative affect (NA) schedules (PANAS) [53] model, using an on-line form of the general PANAS questionnaire [53] which consists of two 10 questions sets, each to access the PA and NA respectively. Participants rated their general feelings in a 5-point intensity scale using questions like "Do you feel in general...?" (e.g. active, afraid See [53]). PANAS is calculated by summing the ratings of all 10 questions for PA and NA respectively, resulting in values between 10 and 50.

The distribution of the Big-Five personality traits, PA and NA, over (i) the 37 participants that took part in the long videos experiment, (ii) the 17 participants of the individual setting, and (iii) the 20 participants of the group setting, are presented in Figure 3. The difference of distribution of ratings, for each of the seven dimensions of personality and PANAS, between the participants of individual and group settings, is not significant ($p > 0.1$ according to a two sample t-test for every dimension).

## 4 Data Analysis

In this section, we present a detailed analysis of the data gathered in both experiments.

### 4.1 Self-Assessment vs External Annotation

The external annotations were validated by assessing the inter-annotator agreement. For this, the annotations corresponding to each participant performed by every annotator were mapped to the $[0, 1]$ range, where 0 corresponds to low and 1 to high valence(arousal), then the Cronbach's $\alpha$ [54] statistic among annotators, commonly used for agreement assessment on continuous scales [7], was calculated. Mean Cronbach's $\alpha$s over all participants of 0.98 for valence and 0.96 for arousal were obtained, which indicates a very strong inter annotator reliability for both dimensions.

With the objective to test at what degree, the affective state of participants assessed through self-assessment, is represented by the external annotations, a comparison between the self-assessment and external annotations of valence and arousal, for the short videos experiment, was performed. For each participant, the Spearman correlation coefficient as well as the $p$-value for the positive correlation test were calculated between the self-assessment scores of each video and the mean external annotation over all the annotators and segments of each video. Assuming independence, the resulting $p$-values were combined to one $p$-value using Fisher's method [55]. For valence, the mean correlation over all participants is $0.44(p < .05)$, and $0.15(p < .05)$ for arousal. These correlations are statistically significant which indicates that the external annotation is a good predictor of the affective state of participants, though for the arousal dimension the correlation is low which shows that it is easier to externally assess valence than arousal.

In Figure 4(a), the distribution of the self-assessment of valence and arousal of all participants for the short videos experiments (16 samples per participant) can be observed. Annotations of each participant have been mapped to the $[-1, 1]$ range. The graph includes circles representing the mean scores, over all participants, of each video. It can be observed that in general valence elicitation worked better than arousal, showing a well defined separation between low and high valence stimuli. Even though the separation of arousal is not as prominent, still there is a difference between low and high arousal stimuli. Figure 4(b) shows the distribution of the external annotations of valence and arousal over the 16 videos of the short videos experiment (94 samples by participant). The mean scores, over all the 20-second clips of each video and all the participants are marked with circles. It can be observed that the data shows a V-shape relating valence and arousal, which is a result of the difficulty of eliciting high-levels of arousal with neutral valence, and high/low levels of valence with low arousal. It can also be observed that in general participants showed the expected affective states (e.g. participants showed higher valence(arousal) with high valence(arousal) content in comparison to low valence(arousal) content), though the difference is not as clear as in self-assessment (Fig. 4(a)).

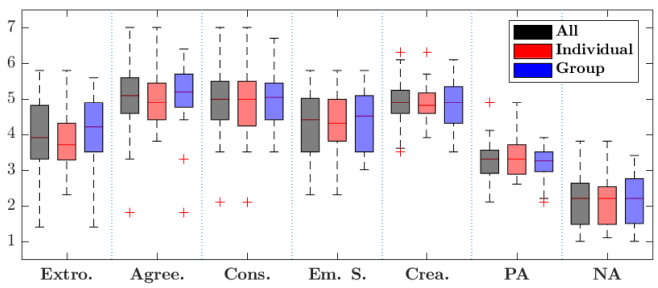



Fig. 3. Distribution of the Big-Five Personality Traits (Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness) and Positive Affect and Negative Affect Schedules (PA and NA) for (i) All, (ii) Individual setting, and (iii) Group setting participants of the Long Videos Experiments. PA and NA are scaled by a 0.1 factor.

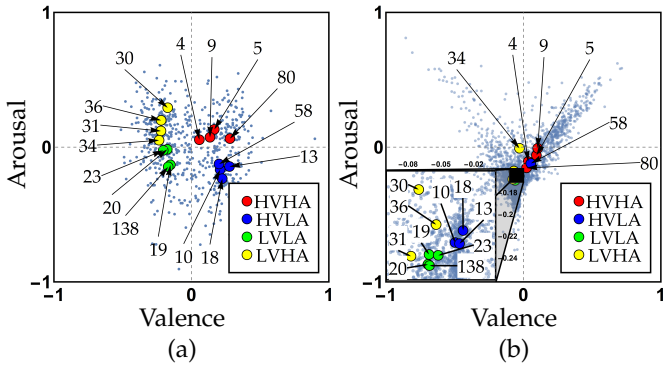



Fig. 4. Distribution of ratings of Valence vs Arousal, for (a) participants' self-assessment of the 16 short videos experiment, and (b) mean external annotations over all annotators for 94 twenty-second segments of the videos of the short videos experiment. Small circles indicate the mean scores over all participants for each of the videos (video ID indicated through arrows). Circles are color coded according to the expected affective response (See Table 2). H, L, V and A, refer to high, low, valence and arousal.

We assessed the effect familiarity has over the affective annotations. For this, we have calculated the Spearman correlation coefficient $\rho$ between the familiarity ratings and the ones of valence ($\rho = 0.127$), arousal ($\rho = 0.014$) and dominance ($\rho = -0.005$). The effect is quite low ($|\rho| \leq 0.127$ for the three cases), which means that familiarity had not a big effect in the participants' reported affective states.

### 4.2 Analysis of Valence and Arousal for Individual and Group Settings

The external annotations of both experiments have been analyzed to test if valence and arousal, expressed by the participants, differed depending on the social context. Two sets of participants were considered. The first set (individual set) corresponds to the 17 participants that took part in the long videos experiment in individual setting, and the second set (group set) corresponds to the 20 participants took part in group setting.

In Fig. 5, the differences in annotations of valence and arousal for the individual set in comparison with the group set for both short and long videos experiments are shown. Fig. 5(a) and (d) show the mean valence and arousal annotations for (i) the individual set (red curve), (ii) the group set (blue curve), and (iii) all participants (black dashed curve), for each of the 340 20s clips. The clips are shown by the video they are part of and ordered according their appearance in the video. In the figure, clips where the difference in the distribution of scores for the group set are significantly lower or higher ($p < 0.05$ according to a two sample t-test) with respect to the one of the individual set are marked with black points and have been shadowed (orange for group scores $<$ individual scores and gray for group scores $>$ individual scores). Fig. 5(b) and (e), show the mean annotations of valence and arousal, for the same sets of participants, of the clips of the short videos experiment, whereas Fig. 5(c) and (f) present the mean annotations for the clips of the long videos experiment. In the (b), (c), (e) and (f) graphs, samples are ordered according to the mean score over all participants (dashed black curve). The clips for which the difference between the distribution of scores from individual and group sets is significant ($p < 0.05$ according to a two sample t-test) are marked with black points.

From Fig. 5(a) and (d) it can be observed that both the high and low areas of the valence and arousal dimensions are covered between all the videos. Comparing the graphs of the short videos experiment (Fig. 5(b) and (e)) with the ones of the long videos experiment (Fig. 5(c) and (f)), it can be observed that in the short videos experiment, where all participants were alone, $21.3\%$ of the clips present significant differences in valence between group and individual participants, and they are concentrated in the low valence region, and $2.1\%$ of the clips present significant differences in arousal. In the long videos experiment, where some participants were in groups, $25.6\%$ of the clips present significant difference of valence between groups and individuals. It is important to note that $48\%$ the clips with significant differences appear in the high valence region (mean valence $> 0$). For arousal, $26.4\%$ of the clips present significant differences between groups and individuals. In Fig. 5(f), where it is observed that in the long videos experiment, group participants showed lower levels of arousal for low arousal clips as well as higher levels of arousal for high arousal clips than individuals.

The Spearman correlation coefficient $\rho$ and the p-value were calculated between the social context label and the mean external annotations for valence and arousal, for the clips of the long videos experiments. The social context label was considered 0 if the participant was in individual setting and 1 if it was in group setting. Significant positive correlation ($\rho = 0.37$, $p < 0.05$) was found between the social context and the mean valence. This significant correlation implies that, in the long videos experiment, participants in group setting showed higher valence than the ones in individual setting. Significant correlation was not found between social context and arousal scores ($p > 0.05$), which suggest that social context does not have a common effect in the arousal expressed by the participants for all clips.

Fig. 5 (c) and (f) show that the scores for clips with low levels of valence(arousal), present a different behavior than the ones with high levels. Therefore, analyses have been independently performed for the low and high valence(arousal) clips of the long videos experiment. For each of the two dimensions (valence and arousal), the clips were sorted based on their score in increasing order, then half of the clips with the lower scores were classified as low class (e.g. low valence) and the other half as high class (e.g. high valence). A two sample t-test of the mean scores of valence(arousal) were performed between the individual and group settings for the clips of low and high classes of valence(arousal). Significant difference was found between individual and group settings for the high valence ($p < 0.001$), low arousal ($p < 0.001$) and high arousal clips ($p < 0.05$), but not for low valence clips ($p = 0.90$). Therefore, social context has an important effect on the valence and arousal expressed by the participants.

### 4.3 Affect, Personality, Mood and Social Context

In Table 5, the Spearman inter-correlations observed between the dimensions of personality, PANAS and social context in the long videos experiment are shown. It also

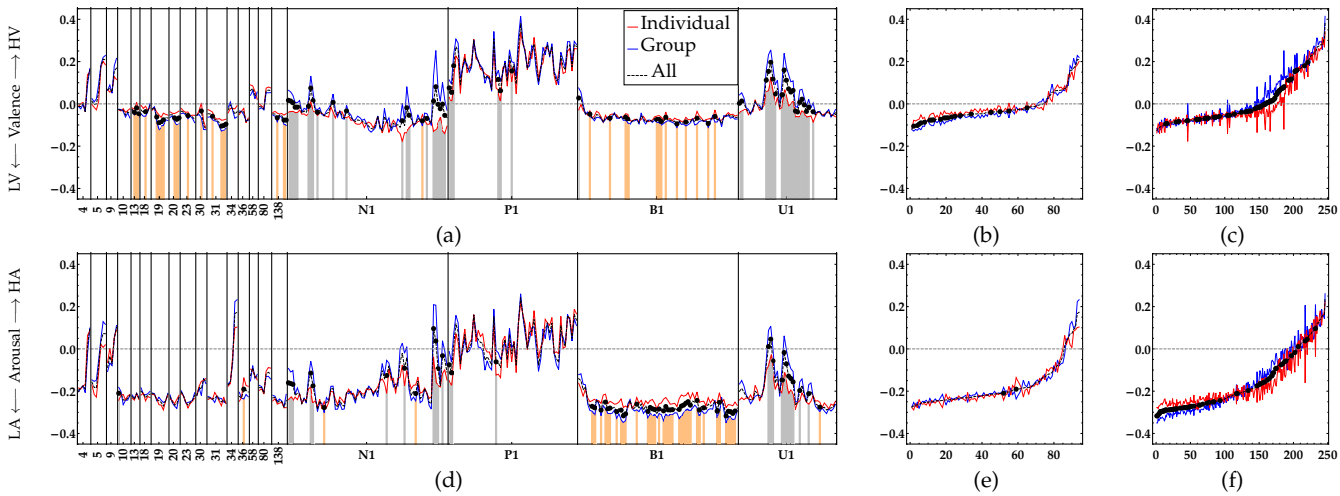



Fig. 5. Mean external annotations of Valence ($V$, upper graphs (a), (b) and (c)) and Arousal ($A$, lower graphs (d), (e), and (f)), over individual participants (red curve), group participants (blue curve) and all participants (dashed black curve), for the videos of ((a) and (d)) both short and long videos experiments (340 segments), ((b) and (e)) the short videos experiment (94 segments), and ((c) and (f)) the long videos experiment (246 segments). Clips where the distribution of scores of individual participants is significantly different than the one of group participants ($p < 0.05$ according to a two sample t-test), are marked with black points. In the case of (a) and (d), video IDs are indicated in the captions. Clips where the distribution of scores of individual participants is significantly higher than the one of group participants ($p < 0.05$), are highlighted in orange. Clips where the distribution of scores of group participants is significantly higher than the one of individual participants are highlighted in gray. In the case of (b), (c), (d) and (f) the horizontal axis represent the number of clips. Origin of valence and arousal (horizontal axis at ($V = 0$) and ($A = 0$)) divides the scale into high-valence (HV: $V > 0$) and low-valence (LV: $V < 0$), and into high-arousal (HA: $A > 0$) and low-arousal (LA: $A < 0$).

TABLE 5
Inter-correlation Between the Dimensions of Personality, PANAS, Social Context in the Long Videos Experiment, and By-participant Mean External Annotations for Valence and Arousal of Short Videos and Long Videos. Significant correlations ($p < 0.05$) are in bold. Ag. Co. E. S., Op. and S. C. refer to Agreeableness, Conscientiousness, Emotional Stability, Openness and Social Context respectively.

| Dims. | Ag. | Co. | E. S. | Op. | PA | NA | S. C. | Valence | | Arousal | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Short | Long | Short | Long |
| Ex. | **0.44*** | 0.09 | 0.21 | 0.13 | 0.32 | **-0.48*** | 0.20 | -0.01 | 0.02 | 0.05 | 0.18 |
| Ag. | - | **0.34*** | 0.14 | 0.24 | **0.43*** | **-0.41*** | 0.18 | -0.21 | 0.00 | 0.13 | 0.21 |
| Co. | - | - | **0.35*** | -0.01 | 0.26 | -0.26 | 0.07 | -0.12 | 0.14 | 0.13 | 0.19 |
| E. S. | - | - | - | 0.24 | -0.12 | **-0.64*** | 0.03 | 0.21 | 0.11 | -0.18 | -0.15 |
| Op. | - | - | - | - | 0.20 | **-0.35*** | -0.04 | 0.23 | 0.13 | 0.06 | 0.02 |
| PA | - | - | - | - | - | -0.06 | -0.03 | -0.03 | 0.16 | 0.30 | **0.61*** |
| NA | - | - | - | - | - | - | -0.01 | -0.28 | -0.02 | -0.12 | 0.04 |

shows the inter-correlations that those dimensions have with the mean external annotations of valence and arousal, of the clips of the short and long videos experiments.

For personality and PANAS, positive significant correlations ($p < 0.05$) were obtained between extraversion and agreeableness, agreeableness and both conscientiousness and PA, and conscientiousness and emotional stability. NA is negatively correlated to all personality and PA dimensions. For social context, significant differences in personality and PANAS distribution between individual and group participants were not obtained, which imply that the group and individual participants have similar distribution of personalities (e.g. individual and group participants have similar levels of extraversion). In general, correlations between personality and PANAS with respect to valence and arousal were not significant, which implies that personality and mood do not necessarily affect the levels of valence and arousal expressed by the participants, with the exception of PA which showed significant positive correlation (0.61) with respect to arousal of the long videos, which indicates that high-PA participants showed higher levels of arousal (they showed more active emotions) than low-PA participants.

## 5 Affect, Personality and PANAS Recognition from Neuro-Physiological Signals

In this section, our baseline methods and results for prediction of affect (valence and arousal), personality, PANAS and social context using neuro-physiological signals are presented. First, the features extracted from the used modalities are described. Next, our method for single modality and fusion of modalities for single-trial classification of affect is presented. Then, our method for single-trial classification of personality traits, PANAS and social context, using single modalities and different schemes for fusion of modalities is presented. Finally, our results are presented and discussed.

### 5.1 EEG, ECG and GSR Features

The neuro-physiological modalities of EEG, ECG and GSR were used to record the participants' implicit responses to affective content. Below, the extracted features from the employed modalities are described. All the features were calculated using the signals recorded during each of the 340 twenty-second clips described in section 3.8.2. Different to other studies that use the concatenation of ECG and GSR as one modality, we study each of them independently to account for the contribution of each one to the recognition task. The summary of features is listed in Table 6.

EEG: Following [1], power spectral density (PSD) features were extracted from the EEG signals. For this, the EEG data was processed using the sampling frequency of 128 Hz. The signals were average-referenced and high-pass filtered with a 2 Hz cut-off frequency. Eye artefacts were removed with a blind source separation technique [56]. By employing the Welch method with windows of 128 samples ($1.0s$),

TABLE 6
Extracted Affective Features for each Modality (feature dimension stated in parenthesis). Computed statistics are: mean, standard deviation (std), skewness, kurtosis of the raw feature over time and % of times the feature value is above/below mean±std.

| Modality | Extracted features |
|---|---|
| **EEG (105)** | 5 bands (theta, slow alpha, alpha, beta and gamma) PSD for each electrode. The spectral power asymmetry between 7 pairs of electrodes in the five bands. |
| **ECG (77)** | Root mean square of the mean squared of IBIs, mean IBI, 60 spectral power in the bands from [0-6] Hz component of the ECG signal, low frequency [0.01,0.08]Hz, medium frequency [0.08,0.15] and hight frequency [0.15,0.5] Hz components of HRV spectral power, HR and HRV stats. |
| **GSR (31)** | Mean skin resistance and mean of derivative, mean differential for negative values only (mean decrease rate during decay time), proportion of negative derivative samples, number of local minima in the GSR signal, average rising time of the GSR signal, spectral power in the [0-2.4] Hz band, zero crossing rate of skin conductance slow response (SCSR) [0-0.2] Hz, zero crossing rate of skin conductance very slow response (SCVSR) [0-0.08] Hz, mean SCSR and SCVSR peak magnitude. |

PSDs, between 3 and 47 Hz, of the signals of every clip were calculated for each of the 14 EEG channels. The obtained PSDs were then averaged over the frequency bands of theta (3-7 Hz), slow alpha (8-10 Hz), alpha (8-13 Hz), beta (14-29 Hz) and gamma (30-47 Hz), and their logarithms were obtained as features. Additionally, the spectral power asymmetry between the 7 pairs of symmetrical electrodes, in the five bands, was calculated. 105 PSD features were obtained (14 channel * 5 bands and 7 symmetrical channels * 5 bands) for every sample (See Table 6).

ECG: Following [57], the heart beats were accurately localized in ECG signals (R-peaks) to calculate the inter beat intervals (IBI). Using IBI values, the heart rate (HR) and heart rate variability (HRV) time series were calculated. Following [6] and [57] 77 features were extracted (See Table 6).

GSR: Following the method of Kim [57], the skin conductance (SC) was calculated from the GSR and then the SC signal was normalized. The normalized signal was low-pass filtered with 0.2 Hz and 0.08 Hz cut-off frequencies to get the low pass (LP) and very low pass (VLP) signals, respectively. Then, the filtered signals were de-trended by removing the continuous piecewise linear trend in the two signals. 31 GSR features employed in [1], [6] were calculated (See Table 6).

## 5.2 Single Trial Classification of Affect in Short and Long Videos

### 5.2.1 Physiological Modalities

For single trial affect (valence and arousal) classification, the features of every modality for each recording session were mapped to the $[-1, 1]$ range in order to avoid the baseline differences that are natural to different recording sessions. This was done for every participant, considering each of the 4 long videos as a recording session and the recordings of the 16 videos of the short videos experiment as a fifth session. For each of the modalities (EEG, ECG and GSR), three scenarios were tested. The first one considers to train and test the system only with the samples of the short videos experiment (94 samples by participant). The second considers only the samples of the long videos experiment (246 samples by participant). The third one considers the combination of the samples of all the videos of both experiments (340 samples by participant), giving in total 9 recognition tasks for every affect dimension.

Leave-one-participant-out cross validation was used, in which, for each affect dimension $j$ label, and for each participant $i$ a Gaussian ($G$) Naïve Bayes (NB) classifier is trained. It assumes independence of the features and is given by:

$$G(f_1, ..., f_n) = \text{argmax}_c\, p(C = c) \prod_{i=1}^{n} p(F_i = f_i | C = c)$$

where $F$ is the set of features and $C$ the classes. $p(F_i = f_i | C = c)$ is estimated by assuming Gaussian distributions of the features and modeling these from the training set. In each step of the cross validation, from the $N$ available participants, the samples of one participant are used as the test set and the samples of the remaining $N - 1$ participants are used as the training set.

For feature selection, Fisher's linear discriminant $J$ [58] defined as $J(f) = \frac{|\mu_1 - \mu_0|}{\sigma_1^2 + \sigma_0^2}$ is calculated for each feature from the training samples. Features are then sorted in decreasing order according to their $J$ value and with a second 10-fold cross-validation over the training set, the optimal $[1 : h]$ most discriminative features are selected. Then, the classifier is trained over all the samples of the training set using the selected features, then it is tested in the test set.

### 5.2.2 Visual Modality

To show the usefulness of the HD video data, we implemented affect recognition using the modality and method of Mou et al [59], [60]. In a nutshell, the method consists of extraction of Volume Quantised Local Zernike Moments (vQLZM) Fisher Vectors [61] as features from the participants' facial HD videos. Dimensionality reduction is applied to the features using PCA preserving the main components that explain 99% of the variance. Finally, classification is performed using a linear-SVM classifier. We refer the reader to [60] for a detailed description of the method.

### 5.2.3 Fusion of Modalities

For each of the three scenarios (short, long and all videos), we implemented decision level fusion of modalities using a linear-SVM as meta-classifier applied over the probabilistic outputs of single modality decisions. We implemented both, physiological modalities fusion (EEG+ECG+GSR), and visual-physiological modalities fusion (Visual+EEG+ECG+GSR).

## 5.3 Classification of Personality, PANAS and Social Context from Short and Long Videos

Connection between physiological signals and personality have been reported in the literature [62]–[64]. Abadi et al [14] inferred Big-Five personality traits and PANAS of 35 participants through the analysis of their implicit responses (EEG, ECG, GSR, and facial landmark trajectories) to 16 short videos, obtaining F1-scores of $70\%$ and $69\%$ for prediction of extraversion and creativity respectively using a linear regression model.

### 5.3.1 Single Modality Classification

For personality traits, PANAS and social context prediction, 7 scenarios have been tested. The different scenarios have been selected to show how the different stimuli as well as their combination perform in the recognition tasks. The

first 4 scenarios (Video-N1, Video-P1, Video-B1 and Video-U1 scenarios) consider only the samples of each of the 4 long videos for prediction. The fifth (Short-videos scenario) considers only the samples of the 16 short videos together. The sixth (Long-videos scenario) considers all the samples of the 4 long videos together. And the seventh (All-videos scenario) considers the samples of all the 20 videos (short and long). The concatenation of the features of all the samples of each scenario and each of the modalities (EEG, ECG and GSR), were associated to the labels of personality traits, PA, NA and social context dimensions. The dimensionality of the feature vector of each scenario is different, for instance the Video N1 scenario with the EEG modality has a feature vector with dimensionality of 7560 features (72 samples $\times$ 105 features) for each participant.

For each scenario and participant, 8 support vector machine (SVM) classifiers with linear kernel [65] were trained, one for each of the 5 personality traits, 2 for mood dimensions of PA and NA and 1 for social context prediction. The labels for personality and mood dimensions are divided into high and low classes using the median value of each personality and mood dimensions as threshold. In the case of social context, if the participant was in a group during the long videos experiment it was considered as positive class and negative if it was in individual configuration. Note that social context prediction was not implemented for the Short-videos scenario simply because it is not applicable.

We use leave-one-participant-out cross-validation, in which, during training, principal components analysis (PCA) [66] is performed over the features of all the participants resulting in a reduction to 36 PCA channels. Next, inspired by [67], channels were selected by clustering them using Pearson correlation coefficient ($\rho$) as distance measure. This is done by ranking the PCA channels according to their Fisher's linear discriminant $J$ calculated for the training set over each channel with respect to the labels. Channels with $J < 0.1$ are discarded. Next, the channel with the highest $J$ is selected. By calculating the $\rho$ coefficient between the selected channel and the remaining channels, redundant channels are removed by discarding channels with $\rho > 0.5$. Then, from the remaining channels the one with the highest $J$ is selected and the process is continued until all the channels are either selected or discarded. With the selected PCA channels, a linear-SVM is trained over the training set and tested over the test set. The regularization parameter $C$ of the linear SVM was empirically set to $0.25$.

#### 5.3.2 Fusion of Modalities

In order to use complementary information from different modalities, decision level fusion of the three modalities (EEG, ECG and GSR) was implemented for each scenario. Following [46], a meta-classification of class labels (M-CLASS) was implemented in which a linear SVM classifier is trained over the probabilistic outputs of the training samples and the training labels. The trained classifier is then used to predict the label of the test sample.

### 5.4 Results and Discussion

In Table 7, the mean F1-scores (mean F1-score for both classes) over all participants, for classification of valence and arousal, using the Gaussian Naïve Bayes classifier, are presented for the physiological signals modalities, and using linear-SVM classifier for visual modality. Three scenarios are included, the first considers only the short videos experiment samples, the second the long videos experiment samples and the third all the samples of both experiments. Results for decision level fusion of both the 3 physiological modalities and of the visual and physiological modalities are also included. Random baseline results (analytically determined) obtained by assigning labels randomly are also included.

TABLE 7
Mean F1-scores (mean F1-score for negative and positive class) over participants for recognition of Valence and Arousal. Bold values indicate whether the F1-score distribution over subjects is significantly higher than $0.5$ according to an independent one-sample t-test ($p < .01$). Analytical results for voting at random are shown.

| Modality | Short | | Long | | All | |
|---|---|---|---|---|---|---|
| | Valence | Arousal | Valence | Arousal | Valence | Arousal |
| EEG | **0.576** | **0.592** | **0.557** | **0.571** | **0.564** | **0.577** |
| GSR | 0.531 | 0.548 | 0.528 | **0.536** | 0.528 | **0.541** |
| ECG | 0.535 | 0.550 | **0.550** | **0.543** | **0.545** | **0.551** |
| Visual (vQLZM-FV) | **0.666** | **0.611** | **0.553** | **0.590** | **0.574** | **0.600** |
| EEG+ECG+GSR | **0.570** | **0.585** | **0.551** | **0.569** | **0.560** | **0.564** |
| Visual+EEG+ECG+GSR | **0.666** | **0.606** | **0.592** | **0.621** | **0.584** | **0.607** |
| Random | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 |

Random levels for all the scenarios for valence and arousal had $0.5$ mean F1-score each. Significant higher than chance ($p < .01$ according to an independent one-sample t-test) F1-scores were obtained for all the scenarios using the EEG modality, for the long videos and all videos scenarios using ECG, and only for arousal recognition in the long videos and all videos scenarios using GSR. In general, arousal recognition got higher performance than valence, except for ECG modality in the long videos experiment. For all scenarios of valence and arousal recognition, EEG got significantly higher performance than ECG and GSR ($p < 0.0001$ for both), resulting in a mean improvement, over the three scenarios, of $2.2\%$ and $3.2\%$ for recognition of valence and arousal over the ECG.

The visual modality shows significant ($p < 0.01$) performance for prediction of valence and arousal in the three scenarios, outperforming the performance of physiological modalities. Decision level fusion of physiological modalities does not improve individual modality results but they are still significantly higher than chance ($p < 0.01$). Fusion of all visual and physiological modalities produces statistically significant ($p < 0.01$) for prediction of valence and arousal for all videos scenarios, outperforming the single modalities in the long and all videos scenarios. Our baseline results show comparable performance with respect to the literature for recognition of valence and arousal [1], [3], [6].

In Table 8, the mean F1-score of the positive and negative classes over all participants for binary classification of personality traits, PANAS and social context is presented. In the table, the seven scenarios described in Sec.5.3.1 are included. We have also implemented the baseline method proposed by Abadi et al [14], based on a linear regression model for predictions using two physiological modalities, namely EEG and physiological signals (ECG+GSR). In [14], they use only short videos and 35 participants. For the sake of comparison,

TABLE 8
Mean F1-score (mean F1-score for negative and positive class) over participants, for personality traits (Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness), PANAS (PA and NA) and social context recognition (number of 20-s segments stated in parenthesis). Bold values indicate whether the F1-score distribution over subjects is significantly higher than $0.5$ according to an independent one-sample t-test ($p < .001$). Results obtained with a baseline method [14], for prediction of personality and PANAS using the short videos experiment are included for comparison. Empirical results for voting at random are also shown.

| Scenario | Modality | Extr. | Agre. | Cons. | Emot. | Open. | PA. | NA. | S. C. |
|---|---|---|---|---|---|---|---|---|---|
| Video N1 (72) | EEG | **0.535** | 0.459 | **0.728** | **0.595** | 0.426 | **0.567** | 0.234 | 0.401 |
| | GSR | **0.675** | **0.699** | 0.284 | 0.405 | 0.459 | 0.431 | 0.327 | **0.644** |
| | ECG | 0.401 | 0.351 | **0.702** | **0.593** | **0.621** | 0.322 | 0.316 | 0.383 |
| Video P1 (58) | EEG | **0.590** | 0.262 | 0.271 | 0.378 | **0.621** | **0.648** | **0.584** | **0.648** |
| | GSR | 0.485 | 0.162 | **0.649** | 0.405 | **0.756** | 0.401 | **0.648** | 0.405 |
| | ECG | 0.431 | 0.405 | **0.619** | **0.619** | 0.431 | **0.648** | **0.584** | 0.405 |
| Video B1 (72) | EEG | **0.675** | **0.619** | **0.644** | 0.324 | 0.135 | 0.401 | **0.745** | 0.449 |
| | GSR | 0.316 | **0.730** | **0.728** | 0.473 | **0.648** | 0.322 | 0.251 | **0.539** |
| | ECG | **0.552** | **0.595** | **0.584** | **0.837** | 0.480 | **0.593** | **0.670** | 0.439 |
| Video U1 (44) | EEG | 0.080 | 0.432 | 0.495 | **0.619** | 0.105 | **0.565** | **0.750** | 0.348 |
| | GSR | 0.431 | **0.675** | 0.348 | **0.730** | **0.560** | 0.485 | **0.598** | 0.401 |
| | ECG | 0.189 | 0.378 | **0.750** | **0.504** | 0.316 | **0.560** | **0.644** | **0.560** |
| Short (94) | EEG | **0.730** | 0.351 | 0.347 | **0.567** | 0.486 | **0.565** | **0.598** | - |
| | GSR | 0.268 | **0.510** | **0.655** | 0.362 | **0.699** | 0.238 | 0.461 | - |
| | ECG | **0.621** | **0.513** | **0.590** | 0.140 | 0.483 | 0.426 | 0.362 | - |
| Long (246) | EEG | **0.756** | 0.405 | 0.271 | **0.539** | 0.378 | 0.485 | **0.619** | **0.528** |
| | GSR | **0.567** | **0.674** | **0.539** | **0.565** | **0.782** | 0.485 | **0.584** | **0.835** |
| | ECG | **0.619** | 0.486 | 0.339 | **0.567** | 0.306 | 0.405 | 0.288 | **0.510** |
| All (340) | EEG | 0.135 | **0.648** | 0.485 | 0.270 | 0.401 | **0.674** | 0.405 | 0.456 |
| | GSR | 0.371 | **0.837** | **0.535** | **0.621** | 0.371 | **0.649** | **0.547** | **0.702** |
| | ECG | 0.485 | **0.567** | 0.449 | 0.189 | **0.648** | 0.459 | **0.590** | **0.728** |
| [14] Abadi et al | EEG | 0.410 | 0.480 | 0.500 | 0.510 | **0.600** | 0.460 | 0.360 | - |
| [14] Abadi et al | ECG+GSR | **0.670** | **0.570** | 0.530 | **0.640** | 0.500 | 0.500 | **0.560** | - |
| Random | - | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 |

we applied their method over the same 37 participants used in this study in the short videos experiment. Empirically estimated baseline results obtained by randomly assigning the labels according to the class ratio of the population are also reported.

Random mean F1-score is $0.5$ for all the scenarios and dimensions (personality traits, PANAS and social context). Different significant ($p < 0.001$) F1-scores are observed for all the scenarios. Single long videos (Video-N1, Video-P1, Video-B1 and Video-U1 scenarios) show to be relevant for the prediction of different personality traits. Consistent significant results over the three modalities are observed for NA prediction in the Video-P1 and Video-U1 scenarios; for agreeableness and consciousness in the Video-B1 scenario; and emotional stability in the Video-U1 scenario. When considering the Short-videos scenario various modalities show contrasting performance. In the Long-videos scenario, consistent significant results are obtained for extroversion, emotional stability and social context. In this scenario, the GSR modality shows the best performance on average for the different dimensions than all other modalities and scenarios with a mean F1-score of $0.623$. In the All-videos scenario, only agreeableness gets consistent performance over each of the modalities.

In comparison with the baseline method [14], using only the short videos with the EEG modality, our method outperforms [14] in prediction of extroversion, emotional stability, PA and NA. It is interesting to note that both methods seem to work complementary to each other. Both methods fail to predict agreeableness and conscientiousness from EEG. Using physiological signals (ECG and GSR), our method outperforms [14] in prediction of conscientiousness and openness using the GSR and in prediction of conscientiousness using ECG. Considering the GSR modality of the Long-videos scenarios, our method outperforms [14] in prediction of agreeableness, conscientiousness, openness and NA.

TABLE 9
Mean F1-score (mean F1-score for negative and positive class) over participants, for recognition of personality traits, PANAS and social context, for fusion of modalities (See 5.3.2). Bold values indicate whether the F1-score distribution over subjects is significantly higher than $0.5$ according to an independent one-sample t-test ($p < .001$). The best performing single modality is also included.

| Scenario | Fusion | Extr. | Agre. | Cons. | Emot. | Open. | PA. | NA. | S. C. |
|---|---|---|---|---|---|---|---|---|---|
| Video N1 | M-CLASS | 0.431 | 0.485 | **0.513** | **0.539** | 0.377 | 0.431 | 0.178 | **0.510** |
| | Best single modality | **0.675** | **0.699** | **0.728** | **0.595** | **0.621** | **0.567** | 0.327 | **0.644** |
| Video P1 | M-CLASS | 0.431 | 0.135 | **0.510** | 0.432 | **0.675** | **0.621** | **0.699** | 0.431 |
| | Best single modality | **0.590** | 0.405 | **0.649** | **0.619** | **0.756** | **0.648** | **0.648** | **0.648** |
| Video B1 | M-CLASS | **0.535** | **0.728** | **0.674** | **0.695** | 0.405 | 0.324 | **0.552** | 0.426 |
| | Best single modality | **0.675** | **0.730** | **0.728** | **0.837** | **0.648** | **0.593** | **0.745** | **0.539** |
| Video U1 | M-CLASS | 0.162 | 0.459 | **0.584** | **0.730** | 0.322 | **0.615** | **0.770** | 0.348 |
| | Best single modality | 0.431 | **0.675** | **0.750** | **0.730** | **0.560** | **0.565** | **0.750** | **0.560** |
| Short | M-CLASS | **0.649** | 0.459 | **0.560** | 0.405 | **0.567** | 0.362 | **0.540** | - |
| | Best single modality | **0.730** | **0.513** | **0.655** | **0.567** | **0.699** | **0.565** | **0.598** | - |
| Long | M-CLASS | **0.648** | **0.510** | 0.268 | **0.513** | **0.535** | 0.449 | **0.699** | **0.725** |
| | Best single modality | **0.756** | **0.674** | **0.539** | **0.567** | **0.782** | 0.485 | **0.619** | **0.835** |
| All | M-CLASS | 0.297 | **0.703** | 0.401 | 0.459 | 0.417 | **0.644** | 0.446 | **0.648** |
| | Best single modality | 0.485 | **0.837** | **0.535** | **0.621** | **0.648** | **0.674** | **0.590** | **0.728** |

Table 9 presents the mean F1-score over all participants for binary classification of personality traits, PANAS and social context, for the decision level fusion scheme described in 5.3.2. The same scenarios as for the single modality experiments are included. The results of the best performing single modalities for each scenario are also included.

We can see from Table 9 that feature level fusion only outperformed the best single modality in a few cases. The difference is only significant for prediction of NA in the Video-P1 and Long-videos scenarios and for prediction of PA in the Video-U1 scenario. In the remaining cases, the weakest modalities seem to undermine the performance of the best modality, but still it is possible to predict conscientiousness and NA in 5 scenarios. It is interesting to note that, though individual long videos do not perform well for social context prediction, using the samples of the 4 long videos experiment together (Long-videos scenario) performs relatively well with mean F1-score of $0.725$. The All-videos scenario which includes samples of both short and long videos does not lead to better performance.

We believe that these results can be improved by the use of different feature extraction and selection methods, such as deep belief networks. We encourage researchers to try and use this challenging dataset.

## 6 Conclusions

In this work, we presented a dataset for multimodal research of affect, personality traits and mood on individuals and groups by means of neuro-physiological signals. We found significant correlations between internal and external affect annotations of valence and arousal, indicating that external

annotation is a good predictor of the affective state of participants. We showed that social context has an important effect on the valence and arousal expressed by the participants, given that group participants showed lower levels of arousal for low arousal clips, and higher levels of arousal for high arousal clips and in general higher valence than when they are alone. PA showed to be significantly correlated with arousal expressed during long videos. For prediction of valence and arousal, EEG was the best physiological modality outperformed only by the visual modality. Decision level fusion of physiological and visual modalities improves individual results. For prediction of personality traits, PANAS and social context, GSR of long videos is the best modality over all dimensions with a mean F1-score of 0.623. Finally, feature level fusion improved the results for NA and PA prediction.

## 7 Acknowledgments

The first author acknowledges support from CONACyT, Mexico, through a scholarship to pursue graduate studies at Queen Mary University of London.

## References

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

[2] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty." *IEEE Trans. on Systems, Man, and Cybernetics, Part A*, vol. 41, no. 6, pp. 1052–1063, 2011.

[3] 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. on Affective Computing*, vol. 6, no. 3, pp. 209–222, July 2015.

[4] S. Koelstra, C. Muehl, and I. Patras, "Eeg analysis for implicit tagging of video data," in *Affective Computing and Intelligent Interaction and Workshops, 2009. 3rd Int'l Conference on*. IEEE, 2009, pp. 1–6.

[5] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang *et al.*, "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3438–3446.

[6] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging." *T. Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[7] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder, "The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent," *IEEE TAC*, vol. 3, no. 1, pp. 5–17, 2012.

[8] A. Dhall, R. Goecke, and T. Gedeon, "Automatic group happiness intensity analysis," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 13–26, Jan 2015.

[9] A. Dhall, J. Joshi, K. Sikka, R. Goecke, and N. Sebe, "The more the merrier: Analysing the affect of a group of people in images," in *Proceedings of the IEEE Int'l Conference on Automatic Face and Gesture Recognition (FG 2015), Ljubljana, Slovenia*, 2015.

[10] X. Huang, A. Dhall, X. Liu, G. Zhao, J. Shi, R. Goecke, and M. Pietikäinen, "Analyzing the affect of a group of people using multi-modal framework," *CoRR*, vol. abs/1610.03640, 2016.

[11] M. Wilson, "Mrc psycholinguistic database: Machine-usable dictionary, version 2.00," *Behavior Research Methods, Instruments, & Computers*, vol. 20, no. 1, pp. 6–10, 1988.

[12] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines," *American Psychologist*, vol. 70, no. 6, pp. 543–556, Sep. 2015.

[13] F. Pianesi, M. Zancanaro, B. Lepri, and A. Cappelletti, "A multimodal annotated corpus of consensus decision making meetings," *Language Resources and Evaluation*, vol. 41, no. 3, pp. 409–429, 2007.

[14] M. Abadi, J. Correa, J. Wache, H. Yang, I. Patras, and N. Sebe, "Inference of personality traits and affect schedule by analysis of spontaneous reactions to affective videos," in *11th IEEE Int'l. Conf. on Automatic Face and Gesture Recog.*, vol. 1, May 2015, pp. 1–8.

[15] J. Wache, R. Subramanian, M. K. Abadi, R.-L. Vieriu, N. Sebe, and S. Winkler, "Implicit User-centric Personality Recognition Based on Physiological Responses to Emotional Videos," in *Proc. of the ACM ICMI*. New York, NY, USA: ACM, 2015, pp. 239–246.

[16] R. Plutchik, "The Nature of Emotions," *American Scientist*, vol. 89, no. 4, pp. 344+, 2001.

[17] P. Ekman and W. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues*. Oxford: Prentice-Hall, 1975.

[18] J. Russell, "A circumplex model of affect," *Jrnl. of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980.

[19] P. Chevalier, J. C. Martin, B. Isableu, and A. Tapus, "Impact of personality on the recognition of emotion expressed via human, virtual, and robotic embodiments," in *24th IEEE Int'l Symp. on Robot and Human Interactive Communication*, 2015, pp. 229–234.

[20] G. Matthews, I. Deary, and M. Whiteman, *Personality Traits*, ser. Personality Traits. Cambridge University Press, 2003.

[21] M. Perugini and L. D. Blas, "Analyzing personality-related adjectives from an etic-emic perspective: The Big Five Marker Scales (BFMS) and the Italian AB5C tax." *BigFive Ass.*, pp. 281–304, 2002.

[22] P. T. Costa and R. R. McCrea, *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, Fla.: Psychological Assessment Resources, 1992.

[23] D. Watson and A. Tellegen, "Toward a consensual structure of mood." *Psychological bulletin*, vol. 98, no. 2, pp. 219–235, Sep. 1985.

[24] D. Watson, L. a. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales." *J Pers Soc Psychol*, vol. 54, no. 6, pp. 1063–70, Jun. 1988.

[25] D. McDuff, R. Kaliouby, T. Senechal, M. Amr, J. Cohn, and R. Picard, "Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected "In-the-Wild"," in *CVPR Workshops*, 2013, pp. 881–888.

[26] S. Mavadati, M. Mahoor, K. Bartlett, P. Trinh, and J. Cohn, "Disfa: A spontaneous facial action intensity database," *IEEE Trans. on Affective Computing*, vol. 4, no. 2, pp. 151–160, 2013.

[27] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the Big-Five personality domains," *Jrnl. of Research in Personality*, vol. 37, no. 6, pp. 504–528, Dec. 2003.

[28] R. Subramanian, J. Wache, M. Abadi, R. Vieriu, S. Winkler, and N. Sebe, "Ascertain: Emotion and personality recognition using commercial sensors," *IEEE TAC*, vol. PP, no. 99, pp. 1–1, 2016.

[29] M. Soleymani, G. Chanel, J. J. M. Kierkels, and T. Pun, *Affective Characterization of Movie Scenes Based on Multimedia Content Analysis and User's Physiological Emotional Responses*, ser. Tenth IEEE Int'l Symposium on Multimedia Multimedia, ISM, 2008, pp. 228–235.

[30] S. Z. Bong, M. Murugappan, and S. Yaacob, *Analysis of Electrocardiogram (ECG) Signals for Human Emotional Stress Classification*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 198–205.

[31] F. Ahmad and O. Olakunle, "Discrete wavelet packet transform for electroencephalogram-based emotion recognition in the valence-arousal space," in *Proceeding of 3rd Int'l Conference on Artificial Intelligence and Computer Science 2015*, 2015, pp. 122–132.

[32] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Emotion, attention, and the startle reflex." *Psycholog. review*, vol. 97, no. 3, p. 377, 1990.

[33] H. Friedman, *Encyclopedia of Mental Health*. Elsevier Science, 2015.

[34] J. Onton and S. Makeig, "High-frequency broadband modulations of electroencephalographic spectra," *Frontiers in human neuroscience*, vol. 3, p. 61, 2009.

[35] A. R. Damasio, T. J. Grabowski, A. Bechara, H. Damasio, L. L. B. Ponto, J. Parvizi, and R. D. Hichwa, "Subcortical and cortical brain activity during the feeling of self-generated emotions," *Nature Neuroscience*, vol. 3, no. 10, pp. 1049–1056, 10 2000.

[36] A. Savran, K. Ciftci, G. Chanel, J. Mota, L. Hong Viet, B. Sankur, L. Akarun, A. Caplier, and M. Rombaut, *Emotion Detection in the Loop from Brain Signals and Facial Images*, ser. Proceedings of the eNTERFACE 2006 Workshop, 2006.

[37] W. Boucsein, *Electrodermal Activity*, ser. Advances in Archaeological and Museum Science. Plenum Press, 1992.

[38] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo, "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," in *Proceedings of the 24th Australian Computer-HI Conference*. New York, NY, USA: ACM, 2012, pp. 420–423.

[39] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," University of Florida, Tech. Rep. A-8, 2008.
[40] M. Liu, D. Fan, X. Zhang, and X. Gong, "Human emotion recognition based on galvanic skin response signal feature selection and svm," in *2016 International Conference on Smart City and Systems Engineering (ICSCSE)*, Nov 2016, pp. 157–160.
[41] J. P. Delaney and D. A. Brodie, "Effects of short-term psychological stress on the time and frequency domains of heart-rate variability," *Percep. and Motor Skills*, vol. 91, no. 2, pp. 515–524, 2000.
[42] R. Homan, J. Herman, and P. Purdy, "Cerebral location of international 1020 system electrode placement," *Electroencephalography and Clinical Neurophysiology*, vol. 66, no. 4, pp. 376 – 382, 1987.
[43] Y. Liu and O. Sourina, "Eeg databases for emotion recognition," in *Int'l Conference on Cyberworlds*, Oct 2013, pp. 302–309.
[44] ——, "Eeg-based valence level recognition for real-time applications." in *CW*. IEEE, 2012, pp. 53–60.
[45] H. Joho, J. Staiano, N. Sebe, and J. M. Jose, "Looking at the Viewer: Analysing Facial Activities to Detect Personal Highlights of Multimedia Contents," *MTAP*, vol. 51, no. 2, pp. 505–523, 2011.
[46] S. Koelstra and I. Patras, "Fusion of facial expressions and EEG for implicit affective tagging," *Image Vision Comput.*, vol. 31, no. 2, pp. 164–174, 2013.
[47] R. Buck, J. Losow, M. Murphy, and P. Costanzo, "Social facilitation and inhibition of emotional expression and communication." *Jrnl. of Personality and Social Psych.*, vol. 63, no. 1, pp. 962 – 968, 1992.
[48] R. McCrae and O. John, "An introduction to the five-factor model and its applications." *J. of Pers.*, vol. 60, no. 2, pp. 175–215, 1992.
[49] S. Koelstra, A. Yazdani, M. Soleymani, C. Mühl, J.-L. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras, "Single Trial Classification of EEG and Peripheral Physiological Signals for Recognition of Emotions Induced by Music Videos," in *Proceedings on Brain Informatics.*, vol. 6334, 2010, pp. 89–100.
[50] M. Soleymani and M. Pantic, "Human-centered implicit tagging: Overview and perspectives." in *SMC*. IEEE, 2012, pp. 3304–3309.
[51] J. Morris, "Observations: SAM: The Self-Assessment Manikin; An Efficient Cross-Cultural Measurement of Emotional Response," *Jrnl. of Advertising Research*, vol. 35, no. 8, pp. 63–38, 1995.
[52] R. Likert, *A Technique for the Measurement of Attitudes*, ser. A Technique for the Measurement of Attitudes. publisher not identified, 1932, no. nos. 136-165.
[53] D. Watson and L. Clark, "The PANAS-X: Manual for the positive and negative affect schedule-expanded form," The University of Iowa, Tech. Rep., 1999.
[54] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.
[55] T. Loughin, "A systematic comparison of methods for combining p-values from independent tests," *Computational Statistics and Data Analysis*, vol. 47, no. 3, pp. 467 – 485, 2004.
[56] G. Gomez-Herrero, K. Rutanen, and K. Egiazarian, "Blind Source Separation by Entropy Rate Minimization," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 153–156, Feb 2010.
[57] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067–2083, Dec 2008.
[58] F. Song, D. Mei, and H. L., "Feature Selection Based on Linear Discriminant Analysis," in *Intelligent System Design and Engineering Application (ISDEA), 2010 Int'l. Conf.*, vol. 1, Oct 2010, pp. 746–749.
[59] W. Mou, H. Gunes, and I. Patras, "Alone versus in-a-group: A comparative analysis of facial affect recognition," in *Proceedings of the 2016 ACM on Multimedia Conference*, ser. MM '16. New York, NY, USA: ACM, 2016, pp. 521–525.
[60] ——, "Automatic recognition of emotions and membership in group videos," in *The IEEE CVPR Workshops*, June 2016.
[61] E. Sariyanidi, H. Gunes, M. Gökmen, and A. Cavallaro, "Local Zernike moment representations for facial affect recognition," in *Proceedings of the British Machine Vision Conference*, 2013.
[62] G. Stenberg, "Personality and the EEG: Arousal and emotional arousability," *Personality and Individual Differences*, vol. 13, no. 1984, pp. 1097–1113, 1992.
[63] B. O. Gilbert, "Physiological and Nonverbal Correlations of Extraversion, Neuroticism, and Psychoticism during Active and Passive Coping," *Personality and individual differences*, vol. 12, no. 12, pp. 1325–1331, 1991.
[64] C. Stough, C. Donaldson, B. Scarlata, and J. Ciorciari, "Psychophysiological correlates of the NEO PI-R Openness, Agreeableness and Conscientiousness: preliminary results," *Int'l Jrnl. of Psychophysiology*, vol. 41, no. 1, pp. 87–91, May 2001.
[65] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. New York, NY, USA: Cambridge University Press, 2000.
[66] I. Jolliffe, *Principal Component Analysis*. Springer Verlag, 1986.
[67] J. Bins and B. A. Draper, "Feature selection from huge feature sets," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2, 2001, pp. 159–165 vol.2.

**Juan Abdon Miranda Correa** received the MSc degree in electronics systems from Tecnológico de Monterrey, Campus Toluca, Mexico, in 2012. He is now working towards the PhD degree at the School of Electronic Engineering and Computer Science, Queen Mary University of London, UK. His research interests include: multimodal affect recognition in human computer interaction, analysis of social interaction in affective multimedia and deep learning.

**Mojtaba Khomami Abadi** is a PhD candidate at the Department of Information Engineering and Computer Science, University of Trento, Italy. Mojtaba is also the CTO of Sensaura Inc., a Canadian startup on real-time and multimodal emotion recognition technologies. His research interests include: user centric affective computing in human computer interaction and affective multimedia analysis.

**Nicu Sebe** received the PhD degree from Leiden University, The Netherlands, in 2001. Currently, he is with the Department of Information Engineering and Computer Science, University of Trento, Italy, where he is leading the research in the areas of multimedia information retrieval and human behavior understanding. He was a general co-chair of ACM Multimedia 2013, and a program chair of ACM Multimedia 2011, ECCV 2016 and ICCV 2017. He is a senior member of the IEEE and ACM and a fellow of IAPR.

**Ioannis Patras** received the PhD degree from the Delft University of Technology, The Netherlands, in 2001. He is a senior lecturer in computer vision in Queen Mary, University of London. He was in the organizing committee of IEEE SMC2004, FGR2008, ICMR2011, ACMMM2013 and was the general chair of WIAMIS2009. His research interests include computer vision, pattern recognition and multimodal HCI. He is a senior member of the IEEE.