Inference of Personality Traits and Affect Schedule by Analysis of Spontaneous Reactions to Affective Videos

Mojtaba Khomami Abadi [*,1,3], Juan Abdón Miranda Correa [*,2], Julia Wache[1], Heng Yang[2],
Ioannis Patras[2], Nicu Sebe[1]

[1] University of Trento, Italy. [2] Queen Mary University of London, UK.
[3] Semantic, Knowledge, and Innovation Lab(SKIL), Telecom Italia.

Abstract— This paper presents a method for inferring the Positive and Negative Affect Schedule (PANAS) and the BigFive personality traits of 35 participants through the analysis of their implicit responses to 16 emotional videos. The employed modalities to record the implicit responses are (i) EEG, (ii) peripheral physiological signals (ECG, GSR), and (iii) facial landmark trajectories. The predictions of personality traits/PANAS are done using linear regression models that are trained independently on each modality. The main findings of this study are that: (i) PANAS and personality traits of individuals can be predicted based on the users' implicit responses to affective video content, (ii) ECG+GSR signals yield 70%±8% F1-score on the distinction between extroverts/introverts, (iii) EEG signals yield 69%±6% F1-score on the distinction between creative/non creative people, and finally (iv) for the prediction of agreeableness, emotional stability, and baseline affective states we achieved significantly higher than chance-level results.

I. INTRODUCTION

In human computer interactions, the emotional state of a user is a great source of information to enrich the experience. For instance, in an e-learning scenario the computer may adapt the content shown to the user depending on how easy (boring) or difficult (stressing) the content is perceived by the user. Recognizing the emotional state of the user to enhance the user experience has been targeted intensively in affective computing research [20]. Some studies used the explicit responses of the users (e.g. interrupting him/her and asking to self-assess his/her emotional state) to access their emotions. However, most of the recent studies [14] try to avoid interruptions and instead analyze the implicit emotional responses of users (e.g. facial expressions, physiological signals) to automatically infer their emotional state. Emotional responses of humans are influenced by some factors (such as mood, baseline affective schedule, personality, temper and memories), that make the emotion recognition tasks more user-specific. However, by learning the effect of the factors on the emotional behaviors, we can generalize the user-specific model to cross-users models.

The objective of this work is to study the relation between these factors (particularly, personality and baseline affective schedule) and the implicit responses of people to affective content. We infer the *BigFive personality traits* [16] and the *Positive and Negative Affect Schedule (PANAS)* [29] by analyzing the features extracted from three modalities, namely (i) EEG signals, (ii) peripheral physiological signals (ECG,

GSR), and (iii) facial landmark trajectories in response to 16 emotional videos.

We performed a mutual information (MI) analysis between different modalities and (i) arousal, (ii) valence, and (iii) dimensions of personality and PANAS. The analysis shows that the implicit responses are informative of the emotional state and the personality/PANAS of individuals. Therefore, the personality traits/PANAS could be revealed at the presence of emotions. We show that (i) emotional states have correlations with personality/PANAS and (ii) emotional state of individuals have normalized mutual information of about 0.5 with different modalities. Therefore, we expect similar levels of mutual information between modalities and personality/PANAS; this is shown to be true especially for peripheral physiological signals.

The main findings can be summarized as follows: *(i)* all three modalities have high mutual information with the dimensions of personality and PANAS, however the relation is not always linear. *(ii)* the peripheral physiological features have relatively higher mutual information with the dimensions of BigFive personality traits and PANAS than the other proposed features; *(iii)* due to strong linear relations between (a) EEG and openness and (b) peripheral physiological signals and extroversion, we achieved remarkably high mean F1-scores (about 70%) on the prediction of high/low extroversion/openness with a simple linear method.

The remainder of the paper is structured as follows: section II summarizes previous research efforts in both (i) personality assessment methods and (ii) emotion measurement through psycho-physiological signals; section III provides an overview of the experimental protocol we followed; section IV describes the data pre-processing and feature extraction steps taken followed by the the mutual information analysis; finally, after reporting experimental results in section V, we discuss them along with the future research directions in section VI.

II. RELATED WORKS

In this section, we review the state of the art on measurement and prediction of affective behavior and personality.

A. Measuring Emotion

Emotions have a large impact on how we experience events in our life. Any behavior and environmental stimulus may have a psychological effect on us and may influence our interpretation of the environment and our consequent

∗ These authors contributed equally to this work.

behavior. Knowing how people feel is helpful in improving interactions both in human-human and in human-computer interaction. Mainly two concepts are used in the literature of affect computing; One classifies emotions in six distinct universal groups [6]: happiness, sadness, anger, fear, surprise and disgust. The other is a dimensional model of emotion that is developed for the continuous measurement of affect. Wundt [30] proposed that emotions could be classified along three dimensions: pleasure, arousal and dominance. Bradley et al. [2] displayed emotions on a two-dimensional plane with the two axes valence (unpleasant-pleasant) and arousal (calm-aroused). Traditional methods to measure emotions are based on questionnaires. In order to detect emotion changes, it is useful to determine baseline levels of positive and negative affect the participants usually experience as it was done by Watson et al. [29] who developed the Positive and Negative Affect Schedule (PANAS).

To avoid the bias that can occur when people rate what they think they are supposed to feel instead of what they actually feel, emotions need to be decoded implicitly. Recent methods use physiological responses [11] or monitor users' facial expressions [8] since both (especially the former) are difficult to control.

Different affective states are correlated to changes in communicative signals such as speech, body language and facial expressions. An extensive review is given in [33]. Many researchers used the implicit responses acquired through psycho-physiological signals to predict the emotional states of humans [14], [12]. Lisetti and Nasoz [14] employed wearable devices to collect the physiological signals such as Galvanic Skin Response (GSR), heart rate (ECG), and skin temperature in order to predict basic emotions. They achieved a maximum 84% emotion recognition accuracy. Abadi et al. [1] measured emotions on the Arousal-Valence dimensions. They compiled a dataset with 30 subjects and used Magnetoencephalogram (MEG), Near Infra-Red (NIR) facial video, Electrooculogram (EOG), Electromyogram (EMG) and ECG responses for 36 emotional movie clips. Koelstra et al. [11] used EMG, EOG, blood volume pulse (BVP), skin temperature, and GSR to predict the emotional state of 32 participants upon watching on music videos. Soleymani et al. [21] created the MAHNOB-HCI multimodal database presenting emotional video clips to participants and collecting physiological signals to predict the emotional state. In this manuscript we take a step forward. By using a subset of videos from [1] and [21] we show that emotional states share high mutual information with personality and PANAS.

B. Personality Assessment

The Big-Five or the five-factor model describes human personality in terms of five dimensions: Extraversion (sociable vs. reserved), Agreeableness (compassionate vs. dispassionate and suspicious), Conscientiousness (dutiful vs. easy-going), Neuroticism or emotional stability (nervous vs. confident), and Openness or Creativity (curious vs. cautious) [16]. The traditional method to measure these personality

dimensions has been through the use of questionnaires or self-reports. Other works used word frequencies in texts, non verbal communication aspects and body language cues for automatic personality recognition. There are few (if any) studies so far covering the connection between physiological signals and personality. The recent review [26] covers most of literature dealing with personality computing. Mairesse et al. [15] used acoustic and lexical features to develop classification, regression and ranking models for personality recognition. Social media are used as well to predict personality, especially with the increasing use of smartphones that can be employed to measure different aspects of communication activities such as calls, instant messages and even frequency of speaking or proximity to other people in their social network. Srivastava et al. [22] presented a novel method for automating personality questionnaire completion utilizing behavioral cues for 50 movie characters, but this was not used in a real-life scenario.

Relationships between personality traits and user responses are mainly reported on Neuroticism and Extroversion ([16], [9], [18]). Stenberg [23] reported relations between personality and arousal in an EEG-based study. According to [23], lower arousal levels are observed for extraverts as compared to introverts, while Neuroticism is associated with high arousal especially for negative valence stimuli. Gilbert [7] used active and passive coping tasks as stimuli and found that heart rate and skin conductance correlate with Extroversion and Neuroticism. Stough et al. [24] found correlations between Openness and Conscientiousness with EEG signals when using photic driving. While previous studies mainly concentrated on finding correlations between implicit responses and personality we employ the implicit responses for the prediction of personality traits. Additionally, to the best of our knowledge, we are the first to use pycho-pysiological responses to predict PANAS.

III. EXPERIMENTAL PROTOCOL AND RATING ANALYSIS

A. Used stimuli and experimental protocol

1) Selected stimuli: Our objective for stimuli selection was to select videos that covered well the arousal and valence (AV) space. For each quadrant of the AV space (High Arousal-High Valence (HAHV), Low Arousal-High Valence (LAHV), Low Arousal-Low Valence (LALV), and High Arousal-Low Valence (HALV)) 3 videos were selected from the 36 videos used in [1]. This selection was made based on the self-assessment of 80 participants. Additionally, one video for each quadrant was selected from the ones used in MAHNOB-HCI [21], giving a set of 16 videos (4 per each quadrant). Selected videos (51s-150s long ($\mu = 86.7$, $\sigma = 27.8$)) are listed in table I. Each video is given an ID that is used to refer to it in the remainder of the paper.

2) Materials and Setup: Experiments were performed in a laboratory environment. Physiological signals were obtained using wearable sensors. EEG was recorded using an Emotiv EPOC Neuroheadset[1] (14 channel {AF3, F7, F3, FC5, T7,

Category	Excerpt's source
HAHV	Airplane (1), When Harry Met Sally (2), Hot Shots (3), Love Actually (4)
LAHV	August Rush (5), Love Actually (6), House of Flying Daggers (7), Mr Beans' Holiday (8)
LALV	Gandhi (9), My girl (10), My Bodyguard (11), The Thin Red Line (12)
HALV	Silent Hill (13), Prestige (14), Pink Flamingos (15), Black Swan (16)

P7, O1, O2, P8, T8, FC6, F4, F8, AF4}, 128 Hz, 14 bit resolution). For ECG and GSR signals recording, two extended Shimmer 2R[2] platforms (12 bit resolution) working at 256 and 128 Hz, were used. A MATLAB[3] based platform running on a PC (Intel Core i7, 3.4 GHz) was used to (i) present the stimuli, (ii) obtain and synchronize the signals, and (iii) get the users' ratings. Subjects were seated approximately at 2 meters from the screen (40-inch,1280 x 1024, 60 Hz) where stimuli were presented at the maximum scale that conserved the original aspect ratio. The sound volume was adjusted for each participant to a comfortable level. Frontal face video was recorded with a JVC GY-HM150E camera.

3) Experimental protocol: 35 healthy participants (12 female), aged between 24 and 40 (mean age 28.85), participated in the experiment.

Preparation: Each participant was informed of the experimental protocol and signed a consent form before she/he was led into the experiment room. The experimenter explained the scales used and how to fill the self-assessment form. Then the sensors were placed and their signals checked. The participant started the experiment once the experimenter left the room.

Experiment pipeline: The recording session started with an initial emotion self-assessment. The 16 videos were presented in a random order in trials consisting of a 5 second baseline recording (fixation cross), the presentation of a short video (see III-A.1), followed by the video emotion self-assessment.

4) Participant self-assessment: At the beginning of the experiment and at the end of each trial, participants performed a self-assessment of their affective state. Self-assessment manikins (SAM) [2] with continuous sliders at the bottom were used to visualize the scales of arousal, valence and dominance. Participants moved the sliders to specify their self-assessment level in a continuous scale. Arousal ranges from *very calm: 1* to *very excited: 9*, valence from *very negative: 1* to *very positive: 9*, and dominance from *overwhelmed with emotions: 1* to *in full control of emotions: 9*. In addition, each participant was also asked to select one or more emotional keywords (Neutral, Disgust, Happiness, Surprise, Anger, Fear, and Sadness) they considered that described their emotional state (1: if chosen, 0: otherwise). The whole experiment including the preparation

steps took 50 minutes on average per person.

B. BigFive Personality and PANAS evaluation

1) Big Five Personality: The Big Five personality traits were measured using the big-five marker scale (BFMS) questionnaire [19]. For each personality trait ten descriptive adjectives were rated on a 7-point Likert scale and the mean was calculated. The distributions of personality measures over all participants are presented in figure 1(a). While for Extroversion and Emotional Stability they are more equally distributed, the average scores for Agreeableness, Conscientiousness and Creativity are more clustered with a higher average close to 5.

2) PANAS: We used the General PANAS questionnaire [29] consisting of 10 questions each to access the positive and the negative affect. The participants filled an online form rating their general feelings on a 5-point intensity scale using questions like "Do you feel in general...?". The positive feelings asked are: active, alert, attentive, determined, enthusiastic, excited, inspired, interested, proud, strong. The negative ones asked are: afraid, scared, nervous, jittery, irritable, hostile, guilty, ashamed, upset, distressed. The resulting positive and negative affect measures are mostly independent as shown in [28]. This allows to investigate both aspects independently. The correlation coefficient is 0.12 which is similar to the ones reported in the literature [28]. PANAS is calculated by summing the values (between 1 and 5) of all 10 questions for PA and NA respectively resulting in values between 10 and 50. The distribution and average for PA and NA is consistent with the literature as well [28]. The mean PA is 32.9 while the mean NA is lower (21.3) as presented in figure 1(b).

C. Affective Rating Analysis

We evaluated the suitability of the presented stimuli in terms of their power to evoke emotions in participants. The mean and standard deviation of participants' self-assessments of arousal and valence for each video is reported in table II. Upon calculating the mean for emotional keywords of each video over participants, the mean values were normalized to sum up to 100 to get the percentage of reported emotional keywords (see table II). According to table II, the chosen stimuli for the four quadrants of the AV space (LALV, HALV, LAHV, and HAHV) generally resulted in the elicitation of
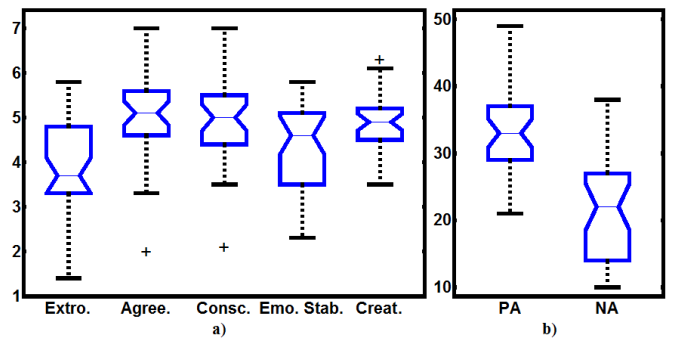

Fig. 1. a) Distribution of the BigFive personality traits. b) Distribution of the average Positive and Negative Affect (PA and NA).

	Video ID	Arousal $\mu \pm \sigma$	Valence $\mu \pm \sigma$	Dominance $\mu \pm \sigma$	%N	%A	%D	%F	%H	%S	%P
HAHV	1	5.5 ± 1.8	5.8 ± 2.0	5.4 ± 2.3	18	8	3	6	**31**	2	**32**
	2	6.0 ± 1.5	6.8 ± 1.3	4.7 ± 1.9	14	1	0	2	**49**	0	**34**
	3	5.5 ± 1.7	6.4 ± 1.3	5.3 ± 2.1	14	2	2	3	**47**	0	**32**
	4	5.4 ± 1.5	7.7 ± 1.0	5.0 ± 1.8	15	0	0	0	**60**	0	**25**
LAHV	5	3.8 ± 1.7	7.0 ± 1.2	6.2 ± 2.2	**44**	0	0	0	**52**	0	4
	6	4.1 ± 1.8	7.7 ± 1.0	5.4 ± 2.1	**32**	0	0	0	**54**	7	7
	7	3.7 ± 1.5	7.3 ± 1.0	5.8 ± 2.0	**27**	0	0	2	**57**	8	8
	8	4.4 ± 1.6	7.0 ± 1.0	6.1 ± 1.9	**33**	0	0	0	**61**	0	6
LALV	9	4.4 ± 1.6	3.9 ± 1.5	6.0 ± 1.8	**32**	26	8	5	3	13	13
	10	5.2 ± 1.6	3.5 ± 1.3	5.0 ± 2.1	24	7	2	2	0	**63**	2
	11	5.0 ± 1.5	3.4 ± 1.4	5.3 ± 2.2	27	**33**	9	5	3	12	11
	12	4.2 ± 1.4	3.6 ± 1.1	5.5 ± 1.9	26	9	5	3	0	**55**	2
HALV	13	6.8 ± 1.4	3.4 ± 1.8	4.5 ± 2.0	17	3	13	**38**	0	1	**28**
	14	5.9 ± 1.5	3.2 ± 1.4	4.7 ± 2.0	11	9	4	20	0	**33**	**23**
	15	5.5 ± 1.4	2.9 ± 2.1	4.0 ± 2.0	7	10	**58**	0	6	2	17
	16	6.5 ± 1.7	3.1 ± 1.4	4.4 ± 1.8	8	4	14	**41**	1	7	25

Dimension	Observed significant($p < 0.05$) correlations
Extroversion	V16 (0.19) D6 (-0.19)
Agreeableness	A2 (-0.08) A8 (-0.42) V3 (-0.04) D16 (-0.04)
Conscientiousness	A7 (0.06) V16 (-0.06)
Emotional Stability	A1 (0.41) A12 (0.13) A14 (0.28) D1 (-0.34)
Creativity (openness)	V6 (0.27) V11 (0.16)
Positive Affect Schedule	A3 (-0.08) A6 (-0.26) V16 (0.04)
Negative Affect Schedule	A15 (-0.22)

the target emotions, and the four quadrants are covered. The relatively lower values of dominance self-assessments over HALV suggest that the participants were more emotionally touched by negative videos.

Among the emotional keywords (adapted from [6]), *happiness* is the only *positive* keyword. We observe *happiness* to be among the dominant emotional keywords chosen for all the HV videos. All HA videos are associated with *surprise* (often as the second dominant keyword), as surprise is characteristic to *excitement*. Interestingly, all the LA videos are labeled with the *neutral* keyword (often as the second dominant keyword), which is due to the lower intensity of emotion in *lower arousing* videos [27].

In LV videos, the underlying negative emotional keywords (sadness/disgust/fear/anger) are often the most dominant reported ones. We observe that the *anger* keyword is only dominant in LA videos. that indicates that the 9th and the 11th video clips involve low intensities of *anger* and evoke *irritation/pity* more than *rage/anger*.

T Wilcoxon signed-rank tests showed that low and high arousal stimuli induced different valence ratings ($p < .005$ and $p < .001$). Similarly, low and high valence stimuli induced different arousal ratings ($p < .0001$ and $p < .0001$). The distribution of the individual ratings per conditions shows a large variance within conditions. This can be explained by between-stimulus and between-participant variations. We investigated the mean inter-correlation of the arousal and valence scales over participants. The mean of the subject-wise inter-correlations between the scales is -0.168. The correlation is significant ($p < .05$) - this is consistent with other studies [11]. Even though the arousal and valence scales are not independent, their negative correlation is quite small implying that participants could differentiate between them.

We measured the Spearman's correlation between the affective ratings of each video provided by the 35 participants versus the BigFive-personality-traits as well as PANAS measures over the 35 participants. The significant observed correlations ($p < .05$) are reported in table III. Previous works established a link between psycho-physiological signals and affective states [17], [10], [11], [21]. Therefore, the obtained correlations between explicit emotional responses (affective self-assessments) and the personality and PANAS dimensions reported in table III, suggest that the implicit emotional responses (i.e. psycho-physiological signals) should also relate to personality and PANAS dimensions. In the next sections, we present a method to predict the personality and PANAS dimensions using a person's implicit responses to emotional videos.

IV. DATA ANALYSIS

We used 3 modalities to record the implicit emotional responses of people: (i) EEG, (ii) peripheral physiological signals (ECG and GSR), and (iii) facial videos. We extracted state of the art *affective features* from different modalities for our analysis. In this section we first describe in detail the extracted features from the employed modalities and then analyze their mutual information with the different affect/personality/PANAS dimensions. To avoid any bias due to different video lengths, all the features are calculated using the responses over the last 50 seconds of the videos.

A. EEG Signal Processing

EEG measures the electrical activity on the scalp. For obtaining features from the EEG signals, the EEG data was processed similarly to [11], using the sampling frequency of 128 Hz. To correct for stimulus-unrelated variations in power over time, the EEG signal from the five seconds before each video was extracted as baseline. Using the Welch method with windows of 128 samples, the frequency power of trials and baseline signals between 3 and 47 Hz was calculated. The baseline power was then subtracted from the trial power, yielding the change of power relative to the pre-stimulus period. These changes of power were averaged over the frequency bands of theta (3-7 Hz), alpha (14-29 Hz), beta (8-13 Hz), and gamma (30-47 Hz). Additionally, the spectral power asymmetry between 7 pairs of electrodes in the four

Modality	Extracted features
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 statistics.
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
EEG (84)	4 bands (theta, alpha, beta, and gamma) spectral power for each electrode. The spectral power asymmetry between 7 pairs of electrodes in the four bands.
Facial Landmark tracks (72)	Statistics concerning horizontal and vertical movement of 12 motion units (MUs) specified in [8].

bands was calculated. The complete set of features is listed in table IV.

B. Peripheral Physiological Signal Processing

We used the methods reported by Kim and Andrè [10] to preprocess the ECG and GSR signals and then extract the features.

Galvanic Skin Response: GSR provides a measure of the electrical resistance of the skin. This resistance varies due to changes in perspiration that are controlled by the sympathetic nervous system (SNS). The changes in GSR are related to the presence of emotions such as stress or surprise while the mean of the GSR signal is related to the level of arousal [13]. In our setup the electrical resistance between two electrodes positioned on the middle phalanges of the middle and index fingers is measured as the GSR signal.

Following [10] we calculated the skin conductance (SC) from GSR and then normalized the SC signal. We low-pass filtered the normalized signal 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, we detrended the filtered signals by removing the continuous piecewise linear trend in the two signals. We calculated 31 GSR features employed in [11], [21] and that are listed in table IV.

Electrocardiogram: The ECG signal was recorded using three electrodes attached to the participant's body. Two of them were placed on the right and left arm crooks and the third one was placed to the left foot as reference. This setup allows precise identification of heart beats. Using the method reported in [10] we accurately localized the heart beats in ECG signals (R-peaks) to calculate the inter beat intervals (IBI). Using IBI values, we calculated the heart rate (HR) and heart rate variability (HRV) time series. Following [21], [10] we extracted 77 features listed in table IV. In this study we use the concatenation of ECG and GSR features as the peripheral physiological features.

C. Facial Video Analysis

We used state of the art methods to initialize and track the facial landmarks and then we extracted statistic measures over 12 motion units (MU) as facial features.

Facial landmark tracking: We extracted the time series of facial landmark location tracks. Before applying the tracking methods, we used the Robust Cascaded Pose Regression (RCPR) [4] with detection model from [32] and the SDM [31] face alignment methods over the first few frames of the facial video. Both of the methods detect the facial landmarks and work in a cascaded way. *SDM* uses only a shape inside the face bounding box as initialization of the face shape (locations of the facial landmarks). In each cascade, based on the calculated Histograms of Oriented Gradient (HoG) features [5] that are calculated. In the surrounding of each landmark, a linear regression is applied. *RCPR* uses several face shape initializations, normalized by the face bounding box. At each cascade, random ferns are used as the primitive regressor for calculating the update. Upon extracting the landmarks using the SDM and RCPR, we validate the correctness by calculating the difference of the locations of their common landmarks. When the difference is smaller than a threshold (set empirically), we use the SDM method to obtain the tracks. Otherwise, the landmarks are set manually. In our experiments, only a few videos, mainly ones in poor lighting conditions needed to be manually checked. The SDM outputs the track of 49 inner facial landmarks using the pixel locations as reference. The landmark detection sample over a frame of a participant's facial video is shown in figure 2.

Processing the facial landmark tracks: To discard the head movement artifact from the facial landmark tracks, we subtracted the track of the nasion (landmark #11) from all the other tracks. Then each track was low-pass filtered with a cut-off frequency of 1Hz. The tracks are used to determine the time series of 12 motion units (MUs) according to [20], [8]. Statistics over the 12 time-series are used as features (see table IV).

D. Mutual Information Analysis

We performed a mutual information analysis between the extracted features from the three modalities versus affect/PANAS/BigFive-personality dimensions (9 in total). Mutual information (MI) between two random variables
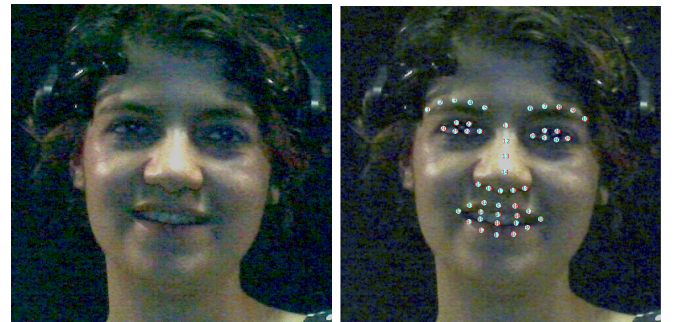

Fig. 2. Left Image: A sample frame of a participant's facial video. Right Image: The output of the SDM facial landmark detection algorithm. The ID of the location of the 49 landmarks are visible under zoom.

measures how much information is known about one of the random variables when the other is known. The function that defines the MI of two random vectors x and y is defined by: $MI(x,y) = \sum_{i,j} p(x_i, y_j) log \frac{p(x_i, y_j)}{p(x_i).p(y_j)}$ where $p(x,y)$ is the joint probability distribution and $p(x)$ and $p(y)$ are the respective marginal probabilities. After calculating the MI between each modality and the affect/PANAS/BigFive-personality dimensions, we calculate the normalized mutual information (NMI) index [25] using the following equation: $NMI(x,y) = \frac{MI(x,y)}{\sqrt{H(x)H(y)}}$ where $H(x)$ and $H(y)$ are the entropies of x and y. We used the MIToolbox [3] to calculate the MI index and entropy values after normalizing x and y to $[0,10]$. Figure 3 presents the normalized 20-bins histograms of the distribution of NMIs for every modality and dimension. The histogram normalization allows a better comparison given the different number of features for every modality. For each normalized histogram, we also calculate the first moment (indicated in red text in figure 3), to summarize the distribution of NMIs. The presented results in figure 3 suggest that (i) the extracted features from different modalities share information with arousal and valence dimensions and hence they contain information about affective state of the participants. The results also suggest that (ii) the features contain information about the participants' personality/PANAS measures. From the two observations we may expect to obtain above-chance prediction of personality and PANAS dimensions, and that may be with the help of affective information included in the extracted features.

V. EXPERIMENTAL SETUP AND RESULTS

In this section we describe our method for the prediction of BigFive-Personality/PANAS based on the extracted features in a *leave-one-subject-out* cross-validation schema.

A. BigFive Personality/PANAS recognition

Each participant watched 16 emotional video clips and for each participant we have five measures for the BigFive personality traits and two measures for PANAS. To this end we extracted the features listed in table IV of 35 participants for each of the 16 emotional videos.

Recognition tasks: We associate all the emotional responses of a participant to his/her personality/PANAS measures and we propose a method that can predict the measures of a new test participant based on the available training data. In total we have seven recognition tasks; five for BigFive-personality-traits and two for PANAS.

Experimental Schema: We use a leave-one-subject-out cross-validation schema to validate our proposed method for solving the *recognition tasks*. Assuming the dataset includes N participants as our samples, in each iteration of the cross validation, we take out one sample as the test sample and use the rest as training samples. We train a linear regression model using the $N-1$ training samples and we predict the measure for the test sample. After completing all the N iterations we dichotomize the prediction and the ground truth values using the median criteria as threshold to divide the samples into high/low classes (e.g. high/low score on extroversion). We then use mean-F1 score of high/low classes to evaluate the quality of the predictions. To more reliably report the performance of our method, we ran the whole cross-validation process 1000 times. In each run, 31 subjects were randomly chosen as samples ($N = 31$) from the 35 available participants. In table V the mean and standard deviation over the obtained results from 1000 runs is presented. The table also includes the random baseline results that are obtained using three methods for the sake of comparison; (i) random voting, (ii) majority class voting, (iii) class distribution voting according to [11]. We also employed a t-test to probe which of the results has a distribution with a significantly ($p < 0.001$) higher than chance level (0.50) mean. The distributions for which the lower bound of the confidence-intervals are more than 0.55 are bold.

Method in Detail: For a certain recognition task (e.g. recognition of extroversion) and a certain modality (e.g. EEG), all the 16 feature vectors in response to 16 emotional videos are taken as samples of a participant. The 16 samples of a participant are associated with the measure of the target dimension (e.g. extroversion). For each participant, the features extracted from each modality in response to 16
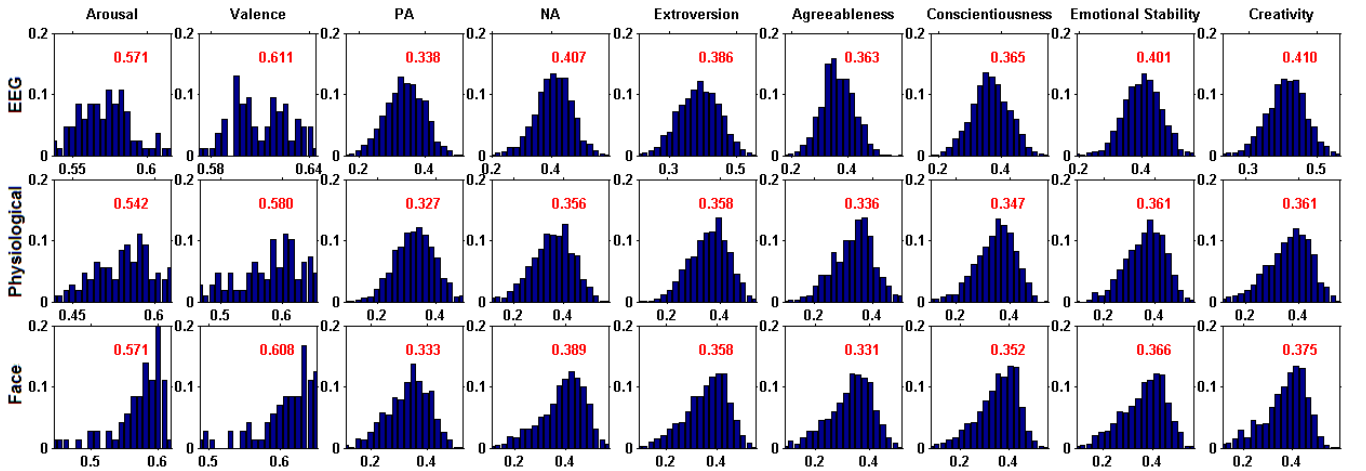

Fig. 3. The normalized histograms of normalized mutual information between each modality and affect/PANAS/BigFive-persoanlity dimensions. The first moment of each distribution is shown in red (best viewed under zoom).

Modality	Ext.	Agr.	Con.	Emo.	Cre.	PA.	NA.
Emotive EEG	0.44± 0.07	**0.60± 0.07**	0.53± 0.08	0.53± 0.07	**0.69± 0.06**	0.38± 0.07	0.49± 0.06
Physiological signals	**0.70± 0.08**	0.50± 0.08	0.53± 0.09	**0.58± 0.08**	0.53± 0.08	**0.60± 0.07**	**0.58± 0.09**
Facial Tracks	0.50± 0.07	**0.58± 0.06**	0.38± 0.09	0.45± 0.07	0.52± 0.08	**0.59± 0.08**	0.48± 0.09
Random Baseline	0.49± 0.09	0.50± 0.09	0.50± 0.09	0.49± 0.09	0.49± 0.09	0.50± 0.09	0.49± 0.09
Majority Baseline	0.50± 0.01	0.50± 0.01	0.50± 0.01	0.50± 0.01	0.50± 0.01	0.50± 0.01	0.50± 0.01
Class Ratio Baseline	0.34± 0.01	0.35± 0.01	0.36± 0.02	0.35± 0.01	0.36± 0.01	0.35± 0.01	0.36± 0.01

video clips are mapped to the range of $[-1,1]$ over the 16 clips. The normalization removes the subjective artifacts and puts the focus of the pattern recognition on differentiating between the responses to different affective videos. During the training, after pooling all the samples ($30 \times 16 = 480$ in total) from the training subjects, we calculate the $z-score$ of features along all the samples. The same parameters of the second normalizations (μ and σ) are used to map the samples of the test subject. We also normalized the scores associated to train/test samples with the parameters of the map to $[-1,1]$. Then we used SVD decomposition to solve the following equation for W_{Tr} (the regression weights):

$$W_{Tr} \times [\mathbf{1} \; D_{Tr}] = S_{Tr} \qquad (1)$$

where D_{Tr} contains the normalized training samples in its columns and S_{Tr} contains the normalized target dimension scores of the train samples in one row vector. We use W_{Tr} to predict P_{Ts}, the prediction of the target dimension of the test subject, using the following equation:

$$W_{Tr} \times [\mathbf{1} \; D_{Ts}] = P_{Ts} \qquad (2)$$

where D_{Ts} includes the normalized test samples in its columns. Table III suggests that the responses to some videos are more useful for the prediction of the target dimension. Therefore, we select a set V of 3 videos that yield the best performance over training samples. Then, we calculate the *median* of the predictions for the videos in V as the estimation of the score for the target dimension of the test subject.

B. Discussion on the Results

Our method for the prediction of different personality/PANAS dimensions is based on a linear regression, therefore it is computationally very cheap but cannot capture nonlinear relations. We observed that different modalities share information (figure 3) with the personality/PANAS dimensions. However not all of the relations are linear. The obtained results presented in table V suggest that the extracted features from peripheral physiological signals have more (strong) linear relation with different dimensions, particularly with extroversion scores. Spectral power features

Dim.Modal.	HAHV				LAHV				LALV				HALV			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Ext. Phys.																
Agr. EEG																
Agr. Face																
Emo.Phys.																
Cre. EEG																
PA. Phys.																
PA. Face																
NA. Phys.																

extracted from EEG responses seem to have strong linear relation with openness. This result is in line with the related exploratory studies [24], [7], [23]. Particularly, (i) Stough et al. [24] found correlations between EEG signals and openness and (ii) Gilbert [7] found that heart rate and skin conductance correlate with Extroversion and Neuroticism. As future work, we will investigate capturing the nonlinear relations between different modalities and target dimension.

C. Chosen videos during the predictions

As mentioned above, each prediction (over a dimension) is based on the predictions over a test person's responses (collected through a modality) to 3 videos, (set V) out of the 16 presented videos. The chosen three videos are the ones that best help the prediction of the dimension. It is interesting to know which videos were *usually* selected for the successful predictions over a dimension. Over the 1000 runs for the prediction of a dimension using a modality, we counted the occurrence of all videos in the chosen set V. Then the percentage of *the times that each video is chosen for the prediction* is calculated and reported in table VI. Since each prediction involves 3 videos, the sum over the entries in each row of the table VI is equal to 300%.

Discussion: For distinguishing between *extroverts/introverts*, videos from all categories were involved. However, low arousal (LA) videos were chosen more often and particularly *August rush (happy)* and *My girl (sad)* were the most effective videos for the prediction of *extroversion*.

For the prediction of *agreeableness*, mainly low arousal videos were selected. The difference between the chosen videos for different modalities (EEG and Facial landmark tracks) for the prediction of agreeableness suggests the presence of complementary information in different modalities and encourages the fusion of information for future extensions of this study.

The important videos in estimating the level of *emotional stability* are selected from all the four quadrants but HALV. HALV videos in our study were rarely chosen for the predictions. The reason may be that the negative videos in HALV (scary/disgusting/stressful videos) are very emotional that touch the majority of population similarly and hence, the

responses of different people to HALV videos are not very useful for the predictions. A support for the last statement is that the HALV videos are shown in table II to be the most emotional ones (with lower values of dominance).

Interestingly, the top videos for the prediction of *openness* are only from HAHV (funny) videos. The observation suggests that the reaction to funny moments in videos is very useful for the prediction of *creativity*.

Positive affect (PA) and negative affect (NA) schedules were mainly estimated through (physiological) responses to positive videos. However in the estimation of the level of PA through facial tracks, *funny videos* (HAHV) had the main role, suggesting that facial expressions to humorous stimuli are distinctive for general positive affect.

VI. Conclusion

This study proposes a method for predicting users' Big-Five personality traits and PANAS of people based on the analysis of their implicit responses to emotional videos. We used 16 emotional videos to evoke emotions in people and recorded the implicit responses through wearable EEG, GSR, and ECG sensors, as well as facial videos. We observed that all the employed modalities share high information with the personality and PANAS dimensions and we showed in some cases a linear model can model well the relation. We tried to capture the linear relations with a linear regressor to predict the correspondent dimensions. The accurate prediction of personality traits and PANAS can later be used (i) to profile people in human-computer interaction and (ii) to develop cross-subject personality/PANAS predictors. Even though we could already show mutual information among constant characteristics (personality traits and General Affect) with changing reactions (EEG, physiological signals and facial expressions), we believe that by using nonlinear regression methods we can obtain even better results. This will be addressed in future work to contribute to better user profiling in human computer interactions.

VII. Acknowledgments

References

[1] 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 Transactions on Affective Computing*, DOI:10.1109/TAFFC.2015.2392932, 2015.

[2] M. Bradley and P. J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(I):49–59, 1994.

[3] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *Journal of Machine Learning Research*, 13(1):27–66, 2012.

[4] X. P. Burgos-Artizzu, P. Perona, and P. Dollár. Robust face landmark estimation under occlusion. In *IEEE Int Conf Computer Vision*, 2013.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.

[6] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *J of personality and social psychology*, 17(2):124, 1971.

[7] B. O. Gilbert. Physiological and Nonverbal Correlates of Extraversion, Neuroticism, and Psychoticism during Active and Passive Coping. *Personality and individual differences*, 12(12):1325–1331, 1991.

[8] 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*, 51(2):505–523, 2011.

[9] E. G. Kehoe, J. M. Toomey, J. H. Balsters, and A. L. W. Bokde. Personality modulates the effects of emotional arousal and valence on brain activation. *Soc Cogn Affect Neurosci*, 7(7):858–70, Oct. 2012.

[10] J. Kim and E. Andre. Emotion recognition based on physiological changes in music listening. *TPAMI,*, 30(12):2067–2083, 2008.

[11] S. Koelstra, C. Mühl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. DEAP: A database for emotion analysis using physiological signals. *IEEE Trans Affective Computing*, 3(1):18–31, 2012.

[12] S. Koelstra and I. Patras. Fusion of facial expressions and eeg for implicit affective tagging. *Image and Vision Computing*, 31(2):164–174, 2013.

[13] P. Lang, M. Bradley, and B. Cuthbert. IAPS: Affective ratings of pictures and instruction manual. Technical report, University of Florida, 2008.

[14] C. L. Lisetti and F. Nasoz. Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. *EURASIP J Applied Sign. Proc.*, 2004(11):1672–1687, 2004.

[15] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *J Artif Intell Res.*, 30:457–500, 2007.

[16] R. R. McCrae and O. P. John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, June 1992.

[17] F. Nasoz, K. Alvarez, C. L. Lisetti, and N. Finkelstein. Emotion Recognition from Physiological Signals for Presence Technologies. *Int. J of Cog., Tech., and Work*, 6(1), 2003.

[18] W. Ng. *Personality and Individual Differences*, 47(1):69–72, 2009.

[19] 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 taxonomy. *Big Five Assessment*, pages 281–304, 2002.

[20] N. Sebe, I. Cohen, and T. S. Huang. Multimodal emotion recognition. In *Handbook of Pattern Recognition and Computer Vision 4*, chapter 1, pages 387–419. 2005.

[21] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Trans Affective Computing*, 3(1):42–55, 2012.

[22] R. Srivastava, J. Feng, S. Roy, T. Sim, and S. Yan. Don't Ask Me What I'm Like, Just Watch and Listen. In *ACM Multimedia*, 2012.

[23] G. Stenberg. Personality and the EEG: Arousal and emotional arousability. *Personality and Individual Differences*, 13(1984):1097–1113, 1992.

[24] C. Stough, C. Donaldson, B. Scarlata, and J. Ciorciari. Psychophysiological correlates of the NEO PI-R Openness, Agreeableness and Conscientiousness: preliminary results. *Int J Psychophysiol.*, 41(1):87–91, May 2001.

[25] A. Strehl and J. Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research*, 3:583–617, 2003.

[26] A. Vinciarelli and G. Mohammadi. A Survey of Personality Computing. *IEEE Trans Affective Computing*, ((to appear)), 2014.

[27] H. L. Wang and L.-F. Cheong. Affective understanding in film. *IEEE Trans CSVT*, 16(6):689–704, 2006.

[28] D. Watson and L. Clark. The PANAS-X: Manual for the positive and negative affect schedule-expanded form. Technical report, The University of Iowa, 1999.

[29] D. Watson, L. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol*, 54(6):1063–70, June 1988.

[30] W. Wundt. *Grundriss der Psychologie (Outlines of Psychology)*. Entgelmann, Leibzig, 1986.

[31] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, 2013.

[32] H. Yang, C. Zou, and I. Patras. Face sketch landmarks localization in the wild. *IEEE Signal Processing Letters*, 2014.

[33] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *TPAMI*, 31(1):39–58, Jan. 2009.