ASCERTAIN: Emotion and Personality Recognition using Commercial Sensors

Ramanathan Subramanian, Senior Member, IEEE, Julia Wache Student Member, IEEE, Mojtaba Khomami Abadi, Student Member, IEEE, Radu L. Vieriu, Member, IEEE, Stefan Winkler, Senior Member, IEEE, Nicu Sebe, Senior Member, IEEE

Abstract—We present ASCERTAIN—an multimodal database for implicit personality and affect recognition using commercial physiological sensors. To our knowledge, ASCERTAIN is the first database to connect personality traits and emotional states via physiological responses. ASCERTAIN contains big-five personality scales and emotional self-ratings of 58 users along with their Electroencephalogram (EEG), Electrocardiogram (ECG), Galvanic Skin Response (GSR) and facial activity data, recorded using off-the-shelf sensors while viewing affective movie clips. We first examine relationships between users’ affective ratings and personality scales in the context of prior observations, and then study linear and non-linear physiological correlates of emotion and personality. Our analysis suggests that the emotion–personality relationship is better captured by non-linear rather than linear statistics. We finally attempt binary emotion and personality trait recognition using physiological features. Experimental results cumulatively confirm that personality differences are better revealed while comparing user responses to emotionally homogeneous videos, and above-chance recognition is achieved for both affective and personality dimensions.

Index Terms—Emotion and Personality recognition, Physiological signals, Multimodal analysis, Commercial sensors

1 INTRODUCTION

Despite rapid advances in Human-computer Interaction (HCI) and relentless endeavors to improve user experience with computer systems, the need for agents to recognize and adapt to the affective state of users has been widely acknowledged. While being a critical component of human behavior, affect is nevertheless a highly subjective phenomenon influenced by a number of contextual and psychological factors including personality.

The personality–affect relationship has been actively studied ever since a correlation between the two was proposed in Eysenck’s personality model [1]. Eysenck posited that Extraversion, the personality dimension that describes a person as either talkative or reserved, is accompanied by low cortical arousal—i.e., extraverts require more external stimulations than introverts. His model also proposed that neurotics, characterized by negative feelings such as depression and anxiety, are more sensitive to external stimulation and become easily upset or nervous due to minor stressors.

Many affective studies have attempted to validate and extend Eysenck’s findings. Some have employed explicit user feedback in the form of affective self-ratings [2], [3], while others have measured implicit user responses such as Electroencephalogram (EEG) activity [4] and heart rate [5] for their analyses. However, few works have investigated affective correlates of traits other than Extraversion and Neuroticism. Conversely, social psychology studies have examined personality mainly via non-verbal social behavioral cues (see [6] for a review), but few works have modeled personality traits based on emotional behavior. Conducting studies to examine the personality–affect relationship is precluded by problems such as subject preparation time, invasiveness of sensing equipment and the paucity of reliable annotators for annotating emotional attributes.

This work builds on [7] and examines the influence of personality differences on users’ affective behavior via the ASCERTAIN database. We utilize ASCERTAIN to (i) understand the relation between emotional attributes and personality traits, and (ii) characterize both via users’ physiological responses. ASCERTAIN contains personality scores and emotional self-ratings of 58 users in addition to their affective physiological responses. More specifically, ASCERTAIN is used to model users’ emotional states and big-five personality traits via heart rate (Electrocardiogram or ECG), galvanic skin response (GSR), EEG and facial activity patterns observed while viewing 36 affective movie clips.

We specifically designed a study with movie scenes as they effectively evoke emotions [8], [9], as typified by genres such as thriller, comedy or horror. Also, different from existing affective databases such as DEAP [10], MAHNOB [11] and DECAF [9], ASCERTAIN comprises data recorded exclusively using commercial sensors to ensure ecological validity and scalability of the employed framework for large-scale profiling applications.

Using the ASCERTAIN data, we first examine correla-

tions among users’ valence (V) and arousal (A) self-ratings and their personality dimensions. We then attempt to isolate physiological correlates of emotion and personality. Our analyses suggest that the relationships among emotional attributes and personality traits are better captured by non-linear rather than linear statistics. Finally, we present single-trial (binary) recognition of A,V and the big-five traits considering physiological responses observed over (a) all, and (b) emotionally homogeneous (e.g., high A, high V) clips. Superior personality recognition is achieved for (b), implying that personality differences are better revealed by comparing responses to emotionally similar stimuli. The salient aspects of ASCERTAIN are:

1. To our knowledge, ASCERTAIN is the first physiological database that facilitates both emotion and personality recognition. In social psychology, personality traits are routinely modeled via questionnaires or social behavioral cues. Instead, this is one of the first works to assess personality traits via affective physiological responses (the only other work to this end is [12]).

2. Different from the DEAP [10], MAHNOB [11] and DECAF [9] databases, we use wearable, off-the-shelf sensors for physiological recordings. This enhances the ecological validity of the ASCERTAIN framework, and above-chance recognition of emotion and personality affirms its utility and promise for commercial applications.

3. We present interesting insights concerning correlations among affective and personality attributes. Our analyses suggest that the emotion–personality relationship is better captured via non-linear statistics. Also, personality differences are better revealed by comparing user responses to emotionally similar videos (or more generally, under similar affect inducement).

From here on, Section 2 reviews related literature to motivate the need for ASCERTAIN, while Section 3 details the materials and methods employed for data compilation. Section 4 presents descriptive statistics, while correlations among users’ affective ratings and personality dimensions are analyzed in Section 5. Section 6 details physiological correlates of emotion and personality, while Section 7 presents recognition experiments. Section 8 discusses the correlation and recognition results, and Section 9 concludes the paper.

2 RELATED WORK

This section reviews related work focusing on (a) multimodal affect recognition, (b) personality assessment and (c) the personality–affect relationship.

2.1 Multimodal affect recognition

As emotions are conveyed by content creators using multiple means (audio, video), and expressed by humans in a number of ways (facial expressions, speech and physiological responses), many affect recognition (AR) methods employ a multimodal framework. Common content-based modalities employed for AR include audio [14]–[16], visual [17]–[19] and audio-visual [20]–[22]. Recent AR methodologies have focused on monitoring user behavior via the use of physiological sensors (see [23] for a review). Emotions induced by music clips are recognized via heart rate, muscle movements, skin conductivity and respiration changes in [24]. Lisetti et al. [25] use GSR, heart rate and temperature signals to recognize emotional states. As part of the HUMAINE project [13], three naturalistic and six induced affective databases containing multimodal data (including physiological signals) are compiled from 8–125 participants. Tavakoli et al. [26] examine the utility of various eye fixation and saccade-based features for valence recognition, while Subramanian et al. [27] correlate user responses with eye movement patterns to discuss the impact of emotions on visual attention and memory.

Koelstra et al. [10] analyze blood volume pressure, respiration rate, skin temperature and Electrooculogram (EOG) patterns for recognizing emotional states induced by 40 music videos. MAHNOB-HCI [11] is a multimodal database containing synchronized face video, speech, eye-gaze and physiological recordings from 27 users. Abadi et al. [9] study Magnetoecephalogram (MEG), Electromyogram (EMG), EOG and ECG responses from users for music and movie clips, and conclude that better emotion elicitation and AR are achieved with movie clips.

2.2 Personality recognition

The big-five or five-factor model [28] describes human personality in terms of five dimensions– Extraversion (sociable vs. reserved), Neuroticism or the degree of emotional stability (nervous vs. confident), Agreeableness (compassionate vs. dispassionate), Conscientiousness (dutiful vs easy-going) and Openness (curious/creative vs. cautious/conservative).

A comprehensive survey of personality computing approaches is presented in [6]. The traditional means to model personality traits are questionnaires or self-reports. Argamon et al. [29] use lexical cues from informal texts for recognizing Extraversion (Ex) and Neuroticism (Neu). Olguin et al. [30] and Alameda-Pineda et al. [31] show that non-verbal behavioral measures acquired using a sociometric badge such as the amount of speech and physical activity, number of face-to-face interactions and physical proximity to other objects is highly correlated with personality. Much work has since employed non-verbal behavioral cues in social settings for personality recognition including [32], where Ex is recognized using speech and social attention cues in round-table meetings, while [33], [34] predict Ex and Neu from proxemic and attention cues in party settings.

Among works that have attempted recognition of all five personality factors, Mairesse et al. [35] use acoustic and lexical features, while Staiano et al. [36] analyze structural features of individuals’ social networks. Srivastava et al. [37] automatically complete personality questionnaires for 50 movie characters utilizing lexical, audio and visual behavioral cues. Brouwer et al. [38] estimate personality traits via physiological measures, which are revealed sub-consciously and more genuinely (less prone to manipulation) than questionnaire answers. In a gaming-based study, they observe a negative correlation between (i) heart rate and Ex, and (ii) skin-conductance and Neu.

2.3 Personality-Affect relationship

The relationship between personality and affect has been extensively examined in social psychology [39], but not in a computational setting. Eysenck’s seminal personality
theory [1] posits that extraverts require more external stimulation than introverts, and that neurotics are aroused more easily. Many studies have since studied the personality–affect relationship by examining explicit or implicit user responses. Personality effects on brain activation related to valence (V) and arousal (A) is investigated in [3], which concludes that Neu correlates negatively with positive V, and positively with A. In an EEG-based study [4], a negative correlation is observed between Ex and A, while a positive correlation is noted between Neu and A especially for negative valence stimuli.

The impact of personality traits on affective user ratings is studied using path analysis in [40]. Feedback scores from 133 students are analyzed in [2] to conclude that neurotics experience positive emotions similar to emotionally stable counterparts in pleasant situations, even though they may experience negative emotions more strongly. Event-related potentials and heart rate changes are studied in [5] to confirm a positive correlation between Neu and A for negative stimuli, while a signal-detection task is used in [41] to suggest that extraverts are generally less aroused than introverts. Brumbaugh et al. [42] examine correlations among the big-five traits, and find Ex and Neu to be associated with increased A while viewing negative videos. Abadi et al. [12] attempt recognition of the big-five traits from affective physiological responses, and our work is most similar to theirs in this respect. Nevertheless, we consider more users and a larger stimulus set in this work (58 users and 36 clips vs 36 users and 16 clips in [12]), and show superior personality trait recognition on comparing physiological responses to emotionally homogeneous clips.

2.4 Spotting the research gap
Examination of related literature reveals that AR methodologies are increasingly becoming user-centric instead of content-centric, suggesting that emotions better manifest via human behavioral cues rather than multimedia content-based (typically audio, visual and speech-based) cues. Nevertheless, the influence of psychological factors such as personality on emotional behavior has hardly been examined, in spite of prior work suggesting that personality affects one’s a) feelings [39], [43], b) emotional perception [3], [4] and c) multimedia preferences [44], [45].

Motivated by the above findings and the lack of publicly available data sets positioned at the intersection of personality and affect, we introduce ASCERTAIN, a multimodal corpus containing physiological recordings of users viewing emotional videos. ASCERTAIN allows for inferring both personality traits and emotional states from physiological signals. We record GSR, EEG, ECG signals using wearable sensors, and facial landmark trajectories (EMO) using a web-camera. In the light of recent technological developments, these signals can be acquired and analyzed instantaneously. Also, Wang and Ji [23] advocate the need for less-intrusive sensors to elicit natural emotional behavior from users. Use of wearable sensors is critical to ensure the ecological validity, repeatability and scalability of affective computing studies, which are typically conducted in controlled lab conditions and with small user groups.

Table 1 presents an overview of publicly available user-centric AR datasets. Apart from being one of the largest datasets in terms of the number of participants and stimuli examined for analysis, ASCERTAIN is also the first database to facilitate study of the personality–affect relationship.

3 ASCERTAIN OVERVIEW
Fig. 1 presents an overview of the ASCERTAIN framework and a summary of the compiled data is provided in Table 2. To study the personality–affect relationship, we recorded users’ physiological responses as they viewed the affective movie clips used in [9]. Additionally, their explicit feedback, in the form of arousal, valence, liking, engagement and familiarity ratings, were obtained on viewing each clip. Finally, personality measures for the big-five dimensions were also compiled using a big-five marker scale (BFMS) questionnaire [46]. We now describe (1) the procedure adopted to compile users’ emotional ratings, personality measures and physiological responses, and (2) the physiological features extracted to measure users’ emotional responses.

3.1 Materials and Methods
Subjects: 58 university students (21 female, mean age = 30) participated in the study. All subjects were fluent in English and were habitual Hollywood movie watchers.

Materials: One PC with two monitors was used for the experiment. One monitor was used for video clip presentation at 1024 × 768 pixel resolution with 60 Hz screen refresh rate, and was placed roughly one meter before the user. The other monitor allowed the experimenter to verify the recorded sensor data. Following informed consent, physiological sensors were positioned on the user’s body as shown in Fig. 2(a). The GSR sensor was tied
to the left wrist, and two electrodes were fixed to the index and middle finger phalanges. Two measuring electrodes for ECG were placed at each arm crook, with the reference electrode placed at the left foot. A single dry-electrode EEG device was placed on the head like a normal headset, with the EEG sensor touching the forehead and the reference electrode clipped to the left ear. EEG data samples were logged using the Lucid Scribe software, and all sensor data were recorded via Bluetooth. A webcam was used to record facial activity. Synchronized data recording and pre-processing were performed using MATLAB Psychtoolbox.

Protocol: Each user performed the experiment in a session lasting about 90 minutes. Viewing of each movie clip is denoted as a trial. After two practice trials involving clips that were not part of the actual study, users watched movie clips randomly shown in two blocks of 18 trials, with a short break in-between to avoid fatigue. In each trial (Fig. 1(b)), a fixation cross was displayed for 4 seconds followed by clip presentation. After viewing each clip, users self-reported their emotional state in the form of affective ratings within a time limit of 30 seconds. They also completed a personality questionnaire after the experiment.

Stimuli: We adopted the 36 movie clips used in [9] for our study. These clips are between 51–127 s long ($\mu = 80, \sigma = 20$), and are shown to be uniformly distributed (9 clips per quadrant) over the arousal-valence (AV) plane.

Affective ratings: For each movie clip, we compiled valence (V) and arousal (A) ratings reflecting the user’s affective impression. A 7-point scale was used with a -3 (very negative) to 3 (very positive) scale for V, and a 0 (very boring) to 6 (very exciting) scale for A. Likewise, ratings concerning engagement (Did not pay attention – Totally attentive), liking (I hated it – I loved it) and familiarity (Never seen it before – Remember it very well) were also acquired. Mean user V,A ratings for the 36 clips are plotted in Fig. 2(b), and are color-coded based on the ground-truth ratings from [9]. Ratings form a ‘C’-shape in the AV plane, consistent with prior affective studies [9], [10].

Personality scores: Participants also completed the big-five marker scale (BFMS) questionnaire [46] which has been used in many personality recognition works [32]–[34]. Scale distributions for the big-five traits are shown in Fig. 2(c). The most and least variance in personality scores are noted for the Extraversion and Openness traits respectively.

3.2 Physiological feature extraction

We extracted physiological features corresponding to each trial over the final 50 seconds of stimulus presentation, owing to two reasons: (1) The clips used in [9] are not

Fig. 1. (a) ASCERTAIN study overview. (b) Timeline for each trial.

Fig. 2. (a) Participant with sensors (EEG, ECG and GSR visible) during the experiment, (b) Mean Arousal-Valence (AV) ratings for the 36 movie clips used in our experiment and (c) Box-plots showing distribution of the big-five personality trait scores for 58 users.
Table 3

<table>
<thead>
<tr>
<th>Modality</th>
<th>Extracted features</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECG (32)</td>
<td>Ten low frequency ([0-2.4] Hz) power spectral densities (PSDs), four very slow response ([0-0.04] Hz) PSDs, IBI, HR and HRV statistics.</td>
</tr>
<tr>
<td>GSR (31)</td>
<td>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 ([0-0.02] Hz), zero crossing rate of skin conductance very slow response ([0-0.08] Hz), mean SCSC and SCVSR peak magnitude.</td>
</tr>
<tr>
<td>Frontal EEG (88)</td>
<td>Average of first derivative, proportion of negative differential samples, mean number of peaks, mean derivative of the inverse channel signal, average number of peaks in the inverse signal, statistics over each of the 8 signal channels provided by the Neurosky software.</td>
</tr>
<tr>
<td>EMO (72)</td>
<td>Statistics concerning horizontal and vertical movement of 12 motion units (MUs) specified in [47].</td>
</tr>
</tbody>
</table>

3.3 Data Quality

A unique aspect of ASCERTAIN with respect to prior affective databases is that physiological signals are recorded using commercial and minimally invasive sensors that allow body movement of participants. However, it is well known that body movements can degrade quality of the recorded data, and such degradation may be difficult to detect using automated methods. Therefore, we plotted the recorded data for each modality and trial, and rated the data quality manually on a scale of 1 (good data)–5 (missing data). For ECG, we evaluated the raw signal from each arm as well as the R-peak amplitude. The presence/absence of facial tracks and correctness of the tracked facial locations were noted for EMO. For GSR, we examined the extent of data noise, and rated ECG (i) on the raw signal, (ii) by summarizing the quality of δ (< 4 Hz), θ (4–7 Hz), α (8–15 Hz), β (16–31 Hz) and γ (> 31 Hz) frequency bands, and (iii) on the pre-calculated attention and meditation channels available as part of the EEG data. Plots and tables with explanations on data quality are available with the dataset. Fig. 3 presents an overview of the data quality for the four considered modalities, with the proportion of trials for which the quality varies from 1–5 highlighted. About 70% of the recorded data is good (corresponding to levels 1-3) for all modalities except ECG, with GSR data being the cleanest. Maximum missing data is noted for EEG, reflecting the sensitivity of the EEG device to head movements.

Fig. 3. Bar plot showing proportion of trials for which data quality ranges from best (1) to worst (5).

4 Descriptive Statistics

In this section, we present statistics relating to user self-reports and personality scores.

4.1 Analysis of Self-ratings

As mentioned previously, we selected 36 movie clips such that their emotional ratings were distributed uniformly over the AV plane as per ground-truth ratings in [9], with 9 clips each corresponding to the HAHV (high arousal-high valence), LAHV (low arousal-high valence), LALV (low arousal-low valence) and HALV (high arousal-low valence) quadrants. The targeted affective state was mostly reached during the ASCERTAIN study as shown in Fig. 2(b). A two-sample t-test revealed significantly higher mean A ratings for consistency’s sake, quadrant-wise video labels derived based on ratings from [9] are used in this work.
for HA clips as compared to LA clips ($t(34) = 5.1253, p < 0.0001$). Similarly, mean V ratings for HV and LV clips were significantly different ($t(34) = 17.6613, p < 0.00005$). Overall, emotion elicitation was more consistent for valence as in prior works [9], [10].

We computed agreement among participants’ A, V ratings using the Krippendorff’s alpha metric– agreement for A and V were respectively found to be 0.12 and 0.58, implying more consensus for clip valence as above. We then computed the agreement between the ASCERTAIN and DECAF [9] populations using the Cohen’s Kappa ($\kappa$) measure. To this end, we computed $\kappa$ between ground-truth (GT) labels from [9] and each user’s A, V labels assigned as high/low based on the mean rating– the mean agreement over all users for A and V was found to be 0.24 and 0.73 respectively. We also computed the $\kappa$ measure between GT and the ASCERTAIN population based on the mean A, V rating of all users– here, an agreement of 0.39 was observed for A and 0.61 for V. Overall, these measures suggest that while individual-level differences exist in affective perception of the movie clips, there is moderate to substantial agreement between assessments of the ASCERTAIN and DECAF populations implying that the considered movie clips are effective for emotion elicitation.

Fig. 4 presents box-plots describing the distribution of the arousal (A), valence (V), engagement (E), liking (L) and familiarity (F) user ratings for (i) all, and (ii) quadrant-based videos. Clearly, low-arousal videos are perceived as more ‘neutral’ in terms of A and V, which leads to the ‘C′ shape in Fig. 2(b). All videos are perceived as sufficiently engaging, while HV clips are evidently more liked than LV clips. Also, the presented movie clips were not very conversant to participants, suggesting that the ASCERTAIN findings are overall unlikely to be influenced by familiarity biases.

### 4.2 Affective Ratings vs Personality Scales

To examine relationships between the different user ratings, we computed Pearson correlations among self-reported attributes as shown in Table 4. Since the analysis involves attribute ratings provided by 58 users for each of the 36 clips, we accounted for multiple comparisons by limiting the false discovery rate (FDR) to within 5% using the procedure outlined in [48]. Highlighted numbers denote correlations found to be significant over at least 15 users (25% of the population) adopting the above methodology.

Focusing on significant correlations, A is moderately correlated with E, while V is found to correlate strongly with L mirroring the observations of Koelstra et al. [10]. A moderate and significant correlation is noted between E and L implying that engaging videos are likely to appeal to viewers’ senses, and similarly, between F and L confirming the mere exposure effect observed in [49] attributing liking to familiarity. Nevertheless, different from [10] with music videos where a moderate correlation is noted between A and V ratings, we notice that the A and V dimensions are uncorrelated for the ASCERTAIN study, which again reinforces the utility of movie clips as good control stimuli. To validate our experimental design, we tested for effects of video length on A, V ratings but did not find any.

Table 5 presents Pearson correlations between personality dimensions. Again focusing on significant correlations, moderate and positive correlations are noted between Extraversion (Ex) and Agreeableness (Ag), as well as between Ex and Openness (O)– prior studies have noted that Ex and O are correlated via the sensation seeking construct [50]. Ag is also found to moderately and positively correlate with Emotional Stability (ES) and O. Conversely, weakly negative-but-insignificant correlations are observed between (i) Ex and ES, and (ii) ES and O.

Partial correlations between emotional and personality attributes are tabulated in Table 6. Considering all movie clips, a significant and moderately negative correlation is noted between Ex and E, implying that introverts were more immersed with emotional clips during the movie-watching experience.
task. A few more significant correlates are observed when mean ratings for quadrant-wise (or emotionally similar) videos are considered. Delineating, Ag is negatively correlated with V for HAHV videos, while the negative correlation between Ex and E manifests for high-arousal (HAHV and HALV) stimuli. Also notable is the moderately negative correlation between Ex and L, and also between Conscientiousness and L for HALV movie clips.

We also performed linear regression analyses with user self ratings as predictors and personality attributes as the target variables for the different video sets, and the coefficients of determination/squared correlations ($R^2$) for the different video sets are presented in Table 7. $R^2$ values with the three best predictors along with the predictor names are listed outside parentheses, while squared correlations with the full model are listed within braces. Considering all movie clips, the best linear model is obtained for Ex with V, E and L ratings as predictors. Among the four AV traits, significant squared correlations are observed for the Ex and Ag traits with V,E,L predictors, and for the ES trait with arousal, valence and liking ratings as predictors for HAHV clips. A significant model is also obtained for Openness with V,E,L predictors considering mildly positive HALV clips. Overall, it is easy to observe from the table that (i) there is little difference in the predictive power of the best-three-predictor and full models, and (ii) the linear models have rather limited predictive power, with the best model explaining only 17% of the personality scale variance. Cumulatively, Tables 6 and 7 cumulatively suggest that the relationship between emotional and personality variables is not well modeled using linear statistics, and it is perhaps worthwhile to explore the use of non-linear measures to this end. From here on, given the high degree of correlation between A and E and between the V and L, we will only focus on A and V dimensions in the rest of the paper.

4.3 Mutual Information Analysis

Mutual information (MI) is a popular metric to capture non-linear relationships between two random variables, and measures how much information is known about one variable given the other. Formally, the MI between two random vectors $X = \{x\}$ and $Y = \{y\}$ is defined as:

$$MI(X,Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}$$

where $P(x,y)$ is the joint probability distribution, while $P(x)$ and $P(y)$ are the respective marginal probabilities. We attempted to describe the relationship between emotional ratings and personality scales via the normalized mutual information (NMI) index [51] defined as:

$$NMI(X,Y) = \frac{MI(X,Y)}{\sqrt{H(X)H(Y)}}$$

where $H(X)$ and $H(Y)$ denote entropies of $X$ and $Y$.

NMI with personality scales for arousal and valence ratings are shown in Fig. 5. In contrast to linear correlations, both A and V share a high degree of mutual information with all five personality traits. Considering all movie clips, emotional ratings share slightly higher MI with A than with V. Also, a strictly higher MI measure is noted when emotionally similar clips are considered instead of all clips. Among personality traits, Ex and Conscientiousness (Con) share the most MI with V,A attributes— in contrast, little correlation is observed between Con and A,V in Table 6. Conversely, lowest MI is noted for Openness (O). One notable difference exists between A and V though— higher MI with arousal is noted for high HV clips, while for all personality traits barring Ag, greater MI with valence is observed for LV clips than for HV clips.

5 Personality Measures vs User Ratings

We now examine the relationship between user V,A ratings and personality scales in the context of hypotheses (H1–H3) put forth in the literature. To this end, we determined high/low trait groups (e.g., emotional stable vs neurotic) for each personality dimension by dichotomizing personality measures based on the median score— this generated balanced high and low sets for the Ex and ES traits, and an unbalanced split for the remaining traits, with the most imbalance (33 vs 25) noted for Ag. We then proceeded to analyze the affective ratings for each group.
5.1 **H1: Extraversion vs Arousal and Valence**

The correlation between Extraversion and arousal has been investigated in many studies—EEG measurements [4], signal detection analysis [41] and fMRI [3] have shown lower arousal in extraverts as compared to introverts, consistent with Eysenck’s personality theory. Also, Ex has been found to correlate with positive valence in a number of works [52]. Analyses presented in Table 6 reveal little correlation between Ex and A for all video categories. While two-sample \( t \)-tests confirmed that extraverts and introverts rated high A and low A videos differently \( p < 0.00001 \) in both cases, no differences could be identified between their A ratings excepting that extraverts provided marginally lower ratings for HA clips \( t(56) = -1.4423, p = 0.0774 \), left-tailed). Focusing on V ratings, positive correlation between Ex and V breaks down for HV clips in Table 6. Two-sample \( t \)-tests also failed to reveal any differences. Therefore, statistical analyses weakly support the negative correlation between Ex and A, but do not corroborate the positive correlation between Ex and V.

5.2 **H2: Neuroticism vs Arousal and Valence**

The relationship between Neu and A has also been extensively studied— a positive correlation between Neu and A is revealed through fMRI responses in [3], and EEG analysis [4] corroborates this observation for negative V stimuli. [2] further remarks that neurotics experience negative emotions stronger than emotionally stable persons. In contrast, differing observations have been made regarding the relationship between Neu and V. Negative correlation between Neu and positive V is noted in [3], whereas a positive relationship between the two for low A stimuli is observed in [40]. [2] remarks that the Neu-V relation is moderated by situation— while neurotics may feel less positive in unpleasant situations, they experience positive emotions as strongly as ES counterparts in pleasant conditions.

Negative correlation between Emotional Stability (ES) and A (or positive correlation between Neu and A) is noted only for HALV clips in Table 6. However, post-hoc \( t \)-tests failed to reveal differences between A ratings for the two categories. Also, Table 6 generally suggests a negative correlation between ES and V— \( t \)-test comparisons further revealed marginally lower V ratings provided by ES subjects for LALV clips \( t(16) = -1.3712, p = 0.0946 \), left-tailed). Overall, our data does not support the positive relationship between Neu and A, and suggests a weakly positive correlation between Neu and V.

5.3 **H3: Openness vs Valence and Arousal**

Among the few works to study Openness, [40] notes a positive correlation between Openness (O) and V under low arousal conditions, which is attributed to the intelligence and sensitivity of creative individuals, enabling them to better appreciate subtly emotional stimuli. Table 6 echoes a positive (even if insignificant) correlation between O and V for LA clips but post-hoc \( t \)-tests to compare V ratings of open and closed groups failed to reveal any differences. However, we noted that closed individuals felt somewhat more aroused by HA clips than open individuals \( t(56) = -1.5011, p = 0.0695 \), left-tailed) as shown in Fig. 6(a). Fine-grained analysis via left-tailed \( t \)-tests to compare quadrant-wise ratings again revealed the slightly higher arousal experienced by closed subjects for HAHV clips \( t(16) = -1.3753, p = 0.0940 \), left-tailed). In summary, our data weakly confirms a positive relationship between O and V as noted in [40], but suggests a negative correlation between O and A.

5.4 **Agreeableness and Conscientiousness**

Table 6 shows a negative but insignificant correlation between Ag and A for HALV videos. Comparison of A ratings by agreeable and disagreeable groups revealed marginally lower A for agreeable subjects for HA clips \( t(56) = -1.2964, p = 0.10 \), left-tailed), and subsequent quadrant-wise comparisons attributed this finding to significantly lower A ratings provided by the agreeable group for strongly negative HALV clips \( t(16) = -2.0587, p < 0.01 \), left-tailed). This trend could possibly be attributed to the association of disagreeable persons with negative feelings such as deceit and suspicion. Table 6 also shows a negative correlation between Ag and V for highly positive HAHV clips. \( T \)-test comparisons again revealed that agreeable subjects provided somewhat lower V ratings for HV clips \( t(56) = -1.4285, p = 0.0793 \), left-tailed), and this was particularly true of HAHV clips for which significantly lower ratings were provided by the agreeable group \( t(16) = -2.0878, p < 0.05 \), left-tailed). Conscientiousness scale differences did not influence VA ratings in any way.

6 **Physiological correlates of emotion and personality**

Linear and non-linear analyses presented in the previous sections suggest that correlations between emotional and personality attributes are better revealed while examining user responses to emotionally similar clips. If explicit ratings provided by users are a conscious reflection of their emotional perception, then the analyses employing physiological signals should also reveal similar patterns. We attempt to identify linear and non-linear physiological correlates of emotion and personality considering responses to all and quadrant-specific clips in this section.

6.1 **Linear correlates of Emotion and Personality**

We attempted to discover physiological correlates of emotional and the big-five personality attributes via partial Pearson correlations. Given the large number of extracted physiological features (Table 3) as compared to the population size for this study, we first performed a principal component analysis (PCA) on each feature modality to avoid overfitting, and retained those components that explained 99% of the variance. This gave us 8–9 predictors for each of the considered modalities. Table 8 presents correlations between these principal components, users’ affective ratings and personality scales \( R^2 \) denotes number of significant correlates). For affective dimensions, we determined significant correlates considering mean user VA ratings provided for the 36 clips. We also trained regression models with
the significantly correlating components as predictors of the dependent emotion/personality variable, and the squared correlations ($R^2$) of these models are also tabulated.

Examining Table 8, the relatively few (maximum of 3) number of significant predictors can be attributed to the sparse number of principal components employed for analysis. Considering correlations with A and V, more correlates are observed for A than for V overall. At least one significant correlate is noted for all modalities except GSR. ECG is found to correlate most with A, with one correlate observed for all video types. ECG also has the most number of correlates with V (one significant correlate for LAHV and HALV clips). One EMO correlate is noted for both A and V respectively in the HAHV and HALV quadrants. A solitary EEG correlate is noted for V considering HALV clips.

A larger number of physiological correlates are observed for personality traits as compared to emotional attributes. Across all five video types, the least number of correlates are noted for Agreeableness, while most correlates are noted for Openness. The ECG modality again corresponds the maximum number of correlates with V (one significant correlate for LAHV and LALV clips). One EMO correlate is noted for both A and V respectively in the HAHV and HALV quadrants. A solitary EEG correlate is noted for V considering HALV clips.

TABLE 8
Physiological correlates of emotion and personality attributes. $R^2$ denotes the number of significant feature correlates, while $R^2$ is the coefficient of determination for the significant correlates model with the significant correlates as predictors. Bold values denote linear regression models with a significant $R^2$ statistic.

<table>
<thead>
<tr>
<th>Video Set</th>
<th>Feature</th>
<th>Arousal $R^2$</th>
<th>Valence $R^2$</th>
<th>Extra. $R^2$</th>
<th>Agreeable $R^2$</th>
<th>Conscient $R^2$</th>
<th>Em. Stab. $R^2$</th>
<th>Open $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>ECG</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>GSR</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>EMO</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td>HAHV</td>
<td>ECG</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>GSR</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>EMO</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td>LAHV</td>
<td>ECG</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>GSR</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>EMO</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td>LALV</td>
<td>ECG</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>GSR</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>EMO</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td>HALV</td>
<td>ECG</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>GSR</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
<tr>
<td></td>
<td>EMO</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
<td>1 $R^2$</td>
</tr>
</tbody>
</table>

6.2 Non-linear correlates
To examine non-linear physiological correlates of emotion and personality, we performed a mutual information analysis as previously between extracted features from the four modalities and the said attributes. Given the varying number of features for each modality, we segregated the NMI distribution over all features and the emotion/personality rating using 10-bin histograms. Fig. 7 presents the first moment or the mean of the NMI histogram distribution computed over the different video sets for each emotional/personality attribute.

It is easy to note from Fig. 7 that personality attributes share more MI with the user physiological responses than A and V, similar to the linear analyses. GSR features share maximum MI with A (highest value of 0.73 for LAHV clips), while EMO features share the most MI with V (peak of 0.75 for HALV clips). In contrast, peak MI of 0.81 is noted between ECG features and Ex. For both emotion and personality attributes, at least one of the NMIs observed for quadrant-based videos is higher than the NMI with all movie clips, implying that a fine-grained examination of the relationship between sub-conscious physiological responses and conscious self-ratings is more informative. Focusing on affective attributes, higher MI between ratings and physiological responses is noted for A for all modalities except EMO. Among the four modalities, ECG and EMO respectively share the most and least MI with A, while EMO and EEG share the highest and least MI with V.

Focusing on the big-five personality traits, the highest NMI histogram means over all modalities are observed for Ex and Con followed by ES, Agree and O. This trend is strikingly similar to the pattern of MI between affective ratings and personality scores obtained in Fig. 4.3. Examining sensing modalities, ECG features share the highest variations under such conditions.
MI with all the personality dimensions, while EEG features correspond to the lowest NMI means.

7 Recognition results

We performed binary recognition of both emotional and personality attributes to evaluate if the proposed user-centric framework can effectively achieve both. This section details the experiments and results thereof.

7.1 Emotion recognition

A salient aspect of our work is the exclusive use of commercial sensors for examining users’ physiological behavior. To evaluate if our emotion recognition results are comparable to prior affective works which used laboratory-grade sensors, we followed a procedure identical to the DEAP study [10]. In particular, the most discriminative physiological features were first identified for each modality using Fisher’s linear discriminant with a threshold of 0.3. Features corresponding to each user were then fed to the naive Bayes (NB) and linear SVM classifiers as shown in Table 9. A leave-one-out cross-validation set. GSR features produce the best recognition performance, while EEG features produce the worst recognition performance. Considering individual modalities, EEG features are better for recognizing A as compared to V, while the remaining three achieve better recognition of V. These results are consistent with earlier observations made in [9], [54]. Considering multimodal results, peripheral (ECG+GSR) features perform better than unimodal features for A recognition, while the best multimodal F1-score of 0.71 is obtained for V. Finally, comparing the two employed classifiers, NB achieves better recognition than linear SVM for both A and V.

7.2 Personality recognition

For binary personality trait recognition, we first dichotomized the big-five personality trait scores based on the median as in Section 5. This resulted in an even distribution of high/low A and V, implying a class ratio (and consequently, a baseline F1-score) of 0.5.

Observing Table 9, above-chance emotion recognition is evidently achieved with physiological features extracted using commercial sensors. The obtained F1-scores are superior to DEAP [10], which can possibly be attributed to (1) the use of movie clips, which are found to be better than music videos for emotional inducement as discussed in [9], and (2) to the considerably larger number of subjects employed in this study, which results in a larger training set. GSR features produce the best recognition performance for both A and V, while ECG features produce the worst recognition performance. Considering individual modalities, EEG features are better for recognizing A as compared to V, while the remaining three achieve better recognition of V. These results are consistent with earlier observations made in [9], [54]. Considering multimodal results, peripheral (ECG+GSR) features perform better than unimodal features for A recognition, while the best multimodal F1-score of 0.71 is obtained for V. Finally, comparing the two employed classifiers, NB achieves better recognition than linear SVM for both A and V.
Based on majority voting, baseline F1-score for the $Ex$ and $ES$ traits is 0.33, and 0.34 for $Ag$, 0.35 for $Con$ and 0.36 for $O$. Via class-ratio based voting, a baseline score of 0.5 is achieved for all traits. We performed PCA on each feature modality in an identical fashion to linear correlation analyses prior to classification. A leave one-subject-out cross-validation scheme was used to compute the recognition results. Three classifiers were employed for recognition, i) naive Bayes, ii) linear (Lin) SVM and iii) Radial Basis Function (RBF) SVM. The $C$ (linear and RBF SVM) and $\gamma$ (RBF SVM) parameters were tuned via leave-one-subject-out grid search cross-validation on the training set.

Table 10 presents the recognition results, with the best F1-scores achieved using unimodal and multimodal features respectively denoted in bold and bold italics. For each personality trait and video set, a better-than-chance recognition F1-score ($> 0.5$) is achieved with at least with one of the considered modalities. Considering user physiological responses to all affective videos, the highest and lowest F1-scores are respectively achieved for $ES$ (0.73) and $O$ (0.53) traits—note from Fig. 2(c) that $ES$ has the second-highest variance among the five personality dimensions, while $O$ corresponds to the lowest variance in personality scores. Excepting for the $ES$ trait, higher recognition scores are generally achieved considering user responses to emotionally similar videos, in line with the findings from linear and non-linear correlation analyses.

For all personality traits except $O$, an F1-score higher than 0.6 is achieved for at least some of the video quadrants. Among feature modalities, ECG features produce the best recognition performance across personality traits and video sets, followed by EEG, GSR and EMO. EEG features are found to be optimal for recognizing $Ex$, while ECG features achieve good recognition for the $Ag$, $Con$ and $ES$ traits. EMO and GSR modalities work best for the Openness trait. Focusing on classifiers, RBF SVM produces the best recognition performance for 13 out of 25 (5 personality traits $\times$ 5 video sets) conditions, while linear SVM performs best only for three conditions. Linear classifiers NB and Lin SVM perform best for the $Ex$ trait, while RBF SVM, performs best for the $O$ trait.

Fusion-based recognition is beneficial, and higher recognition scores are generally achieved via multimodal fusion. With user responses acquired for all videos, the highest and least fusion-based F1-scores are achieved for the $ES$ (0.77 with RBF SVM) and $O$ (0.56 with NB) traits respectively. With quadrant-based videos, a maximum F1-score of 0.78 is noted for $Con$ (with linear SVM). NB classifier works best with fusion-based recognition, and produces best performance for the $Ex$ trait achieving optimal recognition for all the five video sets.

8 Discussion

The correlation analyses and recognition results clearly convey two aspects related to personality recognition from physiological data (i) A fine-grained analysis of users' physiological responses to emotionally similar movie clips enables better characterization of personality differences—this reflects in the better linear models obtained for personality traits considering quadrant-specific videos in Table 8, and the generally higher NMI’s for the same in Fig. 7. Furthermore, higher F1-scores are typically obtained when physiological responses to emotionally similar clips are used for personality trait recognition. (ii) The relationship between personality scales and physiological features is better captured via non-linear metrics—considerably high MI is noted between emotional ratings and personality scores as well as between affective physiological responses and personality traits, and this observation is reinforced with RBF-SVM producing the best recognition performance.

Interesting similarities are also evident from the correlation and recognition experiments. The NB and Lin SVM classifiers work best for the $Ex$ and $ES$ personality traits, for which a number of linear correlates can be noted in Table 8. Also, minimum number of linear physiological correlates are noted for the $Ag$ trait, for which linear classifiers do not work well (best recognition is achieved with RBF SVM for all video types except ‘All’ in Table 9). Likewise, no GSR correlate of emotion is observed in Table 8, which reflects in poor emotion recognition of personality traits with linear classifiers using GSR features in Table 9. Also, only some EMO correlates of personality traits are revealed in Table 8, and this modality achieves inferior personality recognition with linear classifiers.

Comparing Tables 7 and Table 9, EEG shares the least MI with all personality traits among the considered modalities, and RBF SVM performs poorly with EEG features (only one best F1 score in 25 conditions). Conversely, GSR shares considerable MI with personality dimensions, and GSR features work best with the RBF SVM classifier in Table 9. Some discrepancies also arise between the correlation and recognition results. For example, among the big-five personality traits, Openness shares the least MI with all feature modalities but has a number of linear physiological correlates. However, optimal recognition for this trait is achieved with RBF SVM, even though the achieved unimodal F1-scores are the lowest for this trait.

It is pertinent to point out some limitations of this study in general. Weak linear correlations are noted between emotional and personality scores in Table 6, and only few physiological correlat es of emotion and personality are observed in Table 8, which can partly be attributed to the low variance for three of the personality dimensions and particularly the Openness trait, as seen in Fig. 2(c). In this context, median-based dichotomization of the personality scores for binary recognition may not be the most appropriate. However, most user-centered affective studies have also demonstrated recognition in a similar manner and on data compiled from small user populations, due to the inherent difficulty in conducting large-scale affective experiments. Overall, the general consistency in the nature of results observed from the correlation and recognition experiments suggest that data artifacts may have only minimally influenced our analyses, and that reliable affect and personality recognition is achievable via the extracted physiological features. Furthermore, we will make the compiled data publicly available for facilitating related research.

Even though not analyzed in this work, the ASCERTAIN database also includes Familiarity and Liking ratings, which could be useful for other research studies. For example, studying the individual and combined in-
fluence of familiarity, liking and personality traits on affective behavior could be relevant and useful information for recommender systems. In particular, personality-aware recommender systems have become more popular and appreciated of late [55], but the fact that personality differences show up even as consumers watch affective video content can enable video recommender systems to effectively learn user profiles over time.

Familiarity and Liking ratings could be also used to replicate and extend related studies. For example, the study presented in [56] notes a connection between familiarity, liking and the amount of smiling while listening to music. Also, Hamlen and Shuell [57] find a positive correlation between liking and familiarity for classical music excerpts, which increases when an associated video is accompanied by audio. Similar effects could be tested with emotional movie clips via ASCERTAIN.

Finally, the importance of using less-intrusive sensors in affective studies has been widely acknowledged [23], [25]. Minimally invasive and wearable sensors enable naturalistic user response, alleviating stress caused by cumbersome clinical/lab-grade equipment. Choosing minimally invasive sensors is especially critical when complex behavioral phenomena such as emotions are the subject of investigation. While most available affective datasets have been compiled using lab equipment [23], ASCERTAIN represents one of the first initiatives to exclusively employ wearable sensors for data collection, which not only enhances its ecological validity, but also repeatability and suitability for large-scale user profiling.

9 Conclusion

We present ASCERTAIN—a new multimodal affective database comprising implicit physiological responses of 58 users collected via commercial and wearable EEG, ECG, GSR sensors, and a webcam while viewing emotional movie clips. Users’ explicit affective ratings and big-five personality trait scores are also made available to examine the impact of personality differences on AR. Among AR datasets, ASCERTAIN is the first to facilitate study of the relationships among physiological, emotional and personality attributes.

The personality–affect relationship is found to be better characterized via non-linear statistics. Consistent results are obtained when physiological features are employed for analyses in lieu of affective ratings. Finally, AR performance superior to prior works employing lab-grade sensors is achieved (possibly because of the larger sample size used in this study), and above-chance personality trait recognition is obtained with all considered modalities. Personality differences are better characterized by analyzing responses to emotionally similar clips, as noted from both correlation and recognition experiments. Finally, RBF SVM achieves best personality trait recognition, further corroborating a non-linear emotion–personality relationship.
We believe that ASCERTAIN will facilitate future AR studies, and spur further examination of the personality–affect relationship. The fact that personality differences are observable from user responses to emotion-wise similar stimuli paves the way for simultaneous emotion and personality profiling. As recent research has shown that AR is also influenced by demographics such as age and gender [58], we will investigate correlates between affective physiological responses and the aforementioned soft-biometrics in future, coupled with a deeper examination on the relationship between personality and affect. We will also investigate how a-priori knowledge of personality can impact the design of user-centered affective studies.

10 Acknowledgement
This study is supported by the Human-Centered Cyberphysical Systems research grant from A*STAR Singapore, the MIUR Cluster Active Ageing at Home project and Sensaura Inc.

References


