Influence of Demographics and Personality Traits on Physiological Responses to Improve Continuous Emotion Annotation in Video Applications

Swarnali Banik
Computer Science and Information Systems
Bits Pilani K K Birla Goa Campus
Zuarinagar, Sancoale,, India
p20210016@goa.bits-pilani.ac.in

Sougata Sen
Computer Science and Information Systems
BITS Pilani, Goa
Zuarinagar, Goa, India
sougatas@goa.bits-pilani.ac.in

Snehanshu Saha
Computer Science and Information Systems
BITS Pilani K K Birla Goa Campus
Goa, India
HappyMonk AI Labs, India
Goa, India
scibase.snehanshu@gmail.com

Surjya Ghosh
Computer Science and Information Systems
BITS Pilani
Goa, Goa, India
surjya.ghosh@gmail.com

Abstract

Opportunistic emotion annotation (i.e., collecting annotations only at the moments of high emotion fluctuations) is becoming a potential solution to ease the burden of continuous emotion annotation during video consumption. While physiological response variations can be used to detect the opportune moments, the challenge is physiological responses vary across user profiles (e.g., demography, personality). In this paper, we investigated the physiological response (heart rate, galvanic skin response) changes under the influence of continuous emotion stimuli by conducting a user study (N=31) as the subjects (with varied personality traits and demographics) watched eight emotional stimuli videos. The collected dataset (≈7500 video segments spanning 10+ hours) highlights several interesting findings (e.g., females, older adults (30+ years), openness personalities exhibit a higher degree of physiological change). Our findings underscore the importance of factoring in user profile details while aiming to reduce the continuous emotion annotation burden using an opportunistic emotion annotation strategy.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; *Ubiquitous and mobile computing systems and tools*; • **Applied computing** → Anthropology.

Keywords

Continuous emotion annotation, Opportunistic emotion annotation, Physiological response, Demographic differences, Gender, Age, Big five, Personality traits

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3720095

ACM Reference Format:
Swarnali Banik, Sougata Sen, Snehanshu Saha, and Surjya Ghosh. 2025. Influence of Demographics and Personality Traits on Physiological Responses to Improve Continuous Emotion Annotation in Video Applications. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3706599.3720095

1 Introduction

Emotion-aware video applications (e.g., online gaming [23], online tutoring [19], online presentation [30]) improve user engagement by moderating the content flow based on continuously inferred fine-grain user emotion. While state-of-the-art machine learning models can infer fine-grain emotions continuously, training these models requires users to continuously provide emotion annotations (typically collected as emotion self-reports during video consumption, e.g., [17, 35, 36]). Since concentrating on two tasks (video consumption and emotion annotation) in parallel demands significant user attention and degrades the viewing experience, efficient approaches are required to reduce the continuous annotation burden.

In the existing literature, the widely accepted methods to collect momentary, fine-grain, continuous emotions annotation are based on using an auxiliary device (e.g., joystick [35], moving pointer [14]), which the subjects use as an input modality to provide their annotations (emotion self-reports) during video consumption. More recently, Park et al. proposed to collect the annotations for every 5-second segment after video consumption (post-interaction) so that the users do not need to concentrate on two tasks (video consumption and emotion annotation) in parallel [31]. However, the post-interaction annotation approaches may suffer from recall bias [21], and the users need to annotate a large number of segments for a sufficiently long video. To overcome these challenges, Banik et al. proposed opportunistic emotion annotation [5], which suggests identifying (in real-time) the video segments that cause a significant change in the physiological responses and collecting self-reports *only* during these moments instead of continuous annotation.

There are multiple challenges in developing opportunistic annotation strategies for video applications. First, training the opportune moment detection model requires substantial training data from individual users, as physiological responses are personalized [5, 7, 12, 28]. So, it may be challenging to deploy such an approach in a real-world scenario. Second, in the case of the opportunistic annotation strategy, the user is probed for self-reporting in certain video segments (by pausing the video), which causes a significant change in physiological response. Pausing the video and collecting annotations could also negatively impact the video consumption experience [15, 32]. While the challenges are significant, the existing literature on the Experience Sampling Method (ESM) [13, 20] highlights that these challenges may be mitigated by delving deep into the profile information of users. For example, Ghosh et al. demonstrated that sharing data among *similar* users makes it possible to reduce the individual data collection overhead and yet estimate self-reports reliably [16]. Similarly, in peer-assisted self-report collection strategies, self-reports from a set of designated peers are collected to obtain the self-reports of a user when the individual self-reports are not available [6, 9]. Therefore, profile details regarding demographics and personality traits can be essential as similar users may exhibit similar behavioral responses under emotional stimuli [10, 11]. However, to the best of our knowledge, the influence of demographic and personality traits on physiological responses to optimize continuous emotion annotation is under-explored.

Research Questions: Drawing on these gaps, we formulate the following research questions. **(RQ1)** How do demographic factors (e.g., age, gender) influence the variability of physiological responses to sustained emotional stimuli? **(RQ2)** How do personality traits (e.g., openness, conscientiousness) modulate physiological response patterns during continuous emotional experiences? The answers to these research questions (**RQ1, RQ2**) allow us to find if users' physiological responses vary across demographics and personality traits. As a result, a simple data aggregation (like leave-one-subject-out) may not be effective for reducing the continuous emotion annotation effort. Therefore, a more nuanced approach is required to train the opportune moment detection models.

We performed a real-world user study (N=31) to investigate the research questions. In specific, we gathered physiological responses (Galvanic Skin Response (GSR), Heart rate (HR)) and continuous emotion annotations (valence and arousal scores on a scale of 1 to 9 as per the Circumplex model of emotion [33]) as the participants watched a set of stimuli videos (Section 2) using the experiment apparatus developed for the study. We quantified the variation in GSR and HR using change point score [24] (a measurement of change in time-series values) between consecutive video segments (of 5-second duration, the duration is based on earlier works [1, 31]). We compared these change point scores using suitable statistical tests across different demographic factors (age, gender) and personality traits (Big Five [22]). The findings highlight the following:

- the physiological responses vary significantly across genders for all types of emotional stimuli. Female participants exhibit a higher amount of physiological response change than male participants (Section 3.1).
- we also observed that age group has a significant effect on the variation in physiological response. Older adults (age 30+ years) tend to have a higher degree of physiological response variation for all types of stimuli videos (Section 3.2).
- personality traits produce significantly different physiological response variations across emotional stimuli. Specifically, persons with dominant personalities (such as openness or conscientiousness) exhibit a higher degree of change in the physiological response than other dominant personalities (Section 4).

In summary, the findings highlight that by considering user profile details (over a one-size-fits-all approach), opportunistic annotation can be optimized to ease the burden of continuous emotion annotation.

2 User Study

The objective of this user study is to collect physiological responses and continuous emotion annotations from the participants during their video consumption. The collected dataset from this study is used to answer the research questions framed earlier. In this section, we discuss the user study including experiment apparatus, study procedure, and data pre-processing (including personality profiling). This study has been approved by our institute's Human Ethics Committee (HEC), and we obtained the IRB approval prior to conducting the user study.

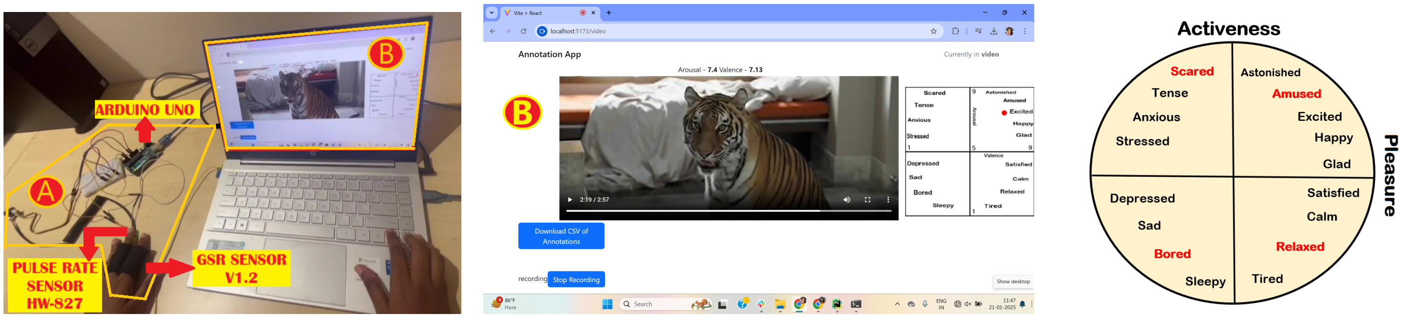
2.1 Experiment Apparatus

The experimental setup comprises two main components, as shown in Fig. 1a. The *first* component (marked as Ⓐ in the figure) consists of a pulse rate sensor (HW-827, World Famous Electronics LLC) and a galvanic skin response (GSR V1.2, Seeed Studio Grove) sensor, both connected to the GPIO pins of an Arduino Uno board. These sensors capture heart rate (HR) and skin conductance (GSR), respectively, at a sampling rate of 10 Hz per sensor. Data from the sensors is transmitted to a connected laptop via a serial port. These sensors were chosen for their ability to provide real-time, accurate data while remaining cost-effective and adaptable for a wide range of applications [29].

The *second* component (marked as Ⓑ in Fig. 1a) is a user interface (UI), referred to as the "Annotate App" (also shown in Fig. 1b), which displays videos and enables users to provide emotion annotations using the keyboard continuously. These annotations are based on the Circumplex model of emotion [33], which represents human emotions on a two-dimensional plane with valence (indicating pleasure) and arousal (indicating activeness), divided into four quadrants (Fig. 1c). When the app is launched, the cursor (shown as a red dot on the Circumplex model of the UI in Fig. 1b) starts at the origin. As the video progresses, users adjust the red dot using the keyboard arrow keys to indicate their experienced emotions in terms of valence and arousal (Fig. 1b). At each timestamp, the cursor's position provides valence and arousal ratings for the video content. When the Annotate App is launched, it simultaneously activates the Arduino to passively record physiological signals (HR and GSR) as users provide emotion annotations during video playback.

2.2 Study Procedure

We recruited 31 participants (15 females, 16 males) aged between 20 and 39 years from our university. We did not apply any specific inclusion criteria while recruiting the participants. Since the study was conducted involving university students, the age range of the



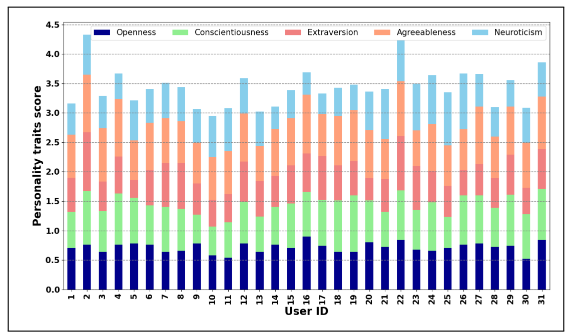
(a) Study setup **(b) The "Annotate App"** **(c) Circumplex model [33]**

Figure 1: The overview of the user study- (a) the study setup, which includes both the sensors (marked as A) and the user interface (UI, marked as B) for video consumption and emotion annotation using keyboard arrows; (b) the UI of the "Annotate App" for video playback and emotion annotation; and (c) the Circumplex model of emotion, which guides the emotion annotation process on a 2D plane (valence and arousal).

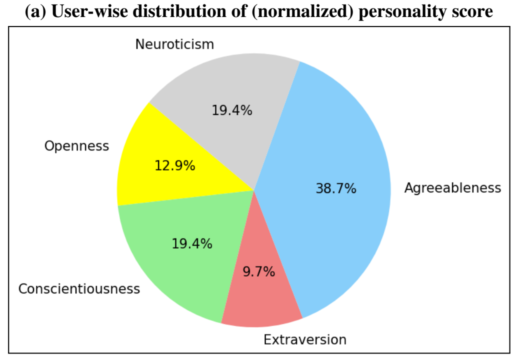
participants turned out between 20 and 39 years. We also collected the personality details from every subject. Each participant filled out the Big Five personality questionnaire [22]. The big five personality traits are measured through a set of 44 questions, with each trait represented by a specific number of questions: Openness (10 questions), Conscientiousness (9 questions), Extraversion (9 questions), Agreeableness (8 questions), and Neuroticism (8 questions). The participant was asked to provide their feedback for each question on a scale of 1 to 5, where one (1) represents "strongly disagree" and five (5) represents "strongly agree". We computed the scores against each dimension (or personality trait) and normalized the score against each dimension by dividing it by the maximum possible score against that dimension. We show the user-wise distribution of (normalized) scores in Fig. 2a. The dimension with the highest normalized score is identified as the participant's dominant personality trait. The distribution of users based on the dominant personality is illustrated in Fig. 2b.

During the study, we recorded physiological signals passively using the sensor setup while the participants watched eight videos embedding a range of emotions based on the Circumplex model (see Fig. 1c). For example, videos 1 and 2 embed amusement (characterized by high valence and high arousal, **HVHA**), while videos 3 and 4 aimed to induce boredom (characterized by low valence and low arousal, **LVLA**). Videos 5 and 6 were selected to elicit relaxation (characterized by high valence and low arousal, **HVLA**), and videos 7 and 8 were intended to provoke fear (characterized by low valence and high arousal, **LVHA**). Each quadrant of the Circumplex model was represented by two videos, ensuring that the full spectrum of valence and arousal was covered during data collection. These videos have also been utilized in similar tasks in previous studies [35].

Each participant viewed the videos in a pre-determined randomized sequence and continuously recorded valence and arousal scores on a scale of 1 to 9 using the keyboard arrow keys. The scores were automatically assigned based on the cursor's position as participants moved the keys. The annotation interface displayed a circumplex plane featuring discrete emotion labels (e.g., amused, scared, bored, relaxed) to guide users in aligning cursor movements with their perceived emotions. This approach enabled users to report the emotions



(a) User-wise distribution of (normalized) personality score



(b) Distribution of users based on dominant personality

Figure 2: Distribution of (a) user-wise normalized personality scores, (the dimension with highest score is identified as the dominant personality of the subject), (b) users across the five dominant personality traits.

they felt, even though the system recorded valence and arousal scores on a 9-point scale. A two-minute blue screen was used between videos to prevent carry-over effects. The participants were briefed on the concepts of valence and arousal and instructed on using the arrow keys to record emotions during video consumption.

Video ID	Emotion	Valence	Arousal	Duration (in sec.)
1	Amusing (**HVHA**)	High	High	185
2				173
3	Boring (**LVLA**)	Low	Low	119
4				160
5	Relaxing (**HVLA**)	High	Low	145
6				147
7	Scary (**LVHA**)	Low	High	197
8				144

Table 1: Details of the stimuli videos embedding four different emotions representing all possible combinations of valence and arousal on the Circumplex plane. H, L indicate high and low respectively, while V, A indicate valence and arousal respectively. The same videos were also used in earlier studies [35] for analyzing continuous emotion annotations.

2.3 Data Pre-processing and Final Dataset

We performed the following data pre-processing tasks on the collected dataset. First, we segmented the physiological signals into fixed-size 5-second windows for each participant and each video, adopting similar approaches outlined in previous studies [1, 31]. Next, we computed the change point score between every two consecutive segments (5-second window) using the RuLSIF algorithm [2, 24]. This change point score measures the extent of change in the physiological responses between two consecutive windows [2]. A high score signifies a major change in physiological signals and suggests an emotional change. Notably, earlier works have used change point score to measure physiological response variation induced by emotional stimuli [1, 5]. We obtained 7415 segments (5-second windows) from all users, equivalent to approximately 10.30 hours of sensor and video data. On average, each user contributed about 239 windows (SD: 1.35). Since all users watched the same set of videos, there was minimal variation in the number of segments per user.

3 RQ1: Demographic Influence on Physiological Response Under Continuous Emotional Stimuli

In this section, we investigate the influence of demographic parameters (gender, age group) on physiological response changes when different emotional stimuli are applied (**RQ1**). We analyze the change point scores to quantify physiological response changes and compare their variations across different demographic factors.

3.1 Influence of Gender on Physiological Response Change

We collate the change point scores for every user as obtained from the physiological response variation captured during the video consumption (Section 2.3). To compare the gender-wise variation in physiological response under continuous emotional stimuli, we perform the following steps. First, we segregate the change-point scores across gender. Next, we checked whether the change-point scores (for male and female) follow normal distribution using the Shapiro-Wilk test [34]. This experiment revealed that the responses did not follow a normal distribution ($p < 0.05$). Since the distribution is not normal and the two groups (male and female) are not paired, we performed the Mann-Whitney U test [25] to evaluate the difference between gender-wise change point scores. The comparison for the gender-wise change point scores across all types of emotional stimuli are shown in Fig. 3a. The mean change point scores for males and females are 2.74 (SD: 2.54) and 3.17 (SD: 2.55), respectively. Also, the medians of change point scores (male and female) are 2.26 and 2.83, respectively. The effect size (Cohen's d) is 0.17. The Mann-Whitney's U test found a significant difference in the gender-wise change point scores (U = 6110129.0, Z = 0.84, p = 0.00).

Next, we performed the same steps for each type of stimuli videos (i.e., **HVHA, HVLA, LVHA, LVLA** as noted in Table 1) independently. The results for each stimuli type are presented in Fig. 3b, 3c, 3d, and 3e respectively. We observed a significant difference in the change point scores across genders for **HVHA** videos. The test statistics are as follows (U = 491848.0, Z = 0.82, p = 0.00). We note that for **HVLA**, **LVHA**, and **LVLA** videos, there is a significant difference. The test statistics are as follows ({**HVLA**: U = 320884.0, Z = 0.93, p = 0.00}; {**LVHA**: U = 435164.50, Z = 0.84, p = 0.00}; {**LVLA**: U = 297257.0, Z = 0.78, p = 0.00}).

In summary, these findings demonstrate that significant differences exist in the physiological responses between male and female participants (females exhibit a high degree of change) for all types of continuous emotional stimuli. Therefore, it may not be optimal to use a generic model (without considering gender) to detect the emotional variances.

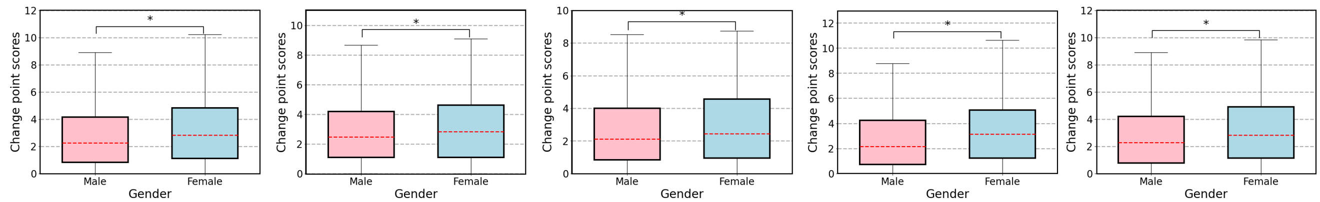
3.2 Influence of Age on Physiological Response Change

We also investigated the influence of age (another demographic factor) on physiological responses under the continuous emotion stimuli. We divided the subjects into two groups (20-29 years and 30-39 years) considering the adulthood impact on emotional changes based on the developmental theory [3, 4, 8].

In this case, also, we used change point score (Section 2.3) as the quantitative measure of physiological response variation. We performed two types of comparison. First, we investigated if there is a significant variation in the change point scores for all types of stimuli videos for different age groups. Later, we compared the stimuli type-wise variation in the change point scores.

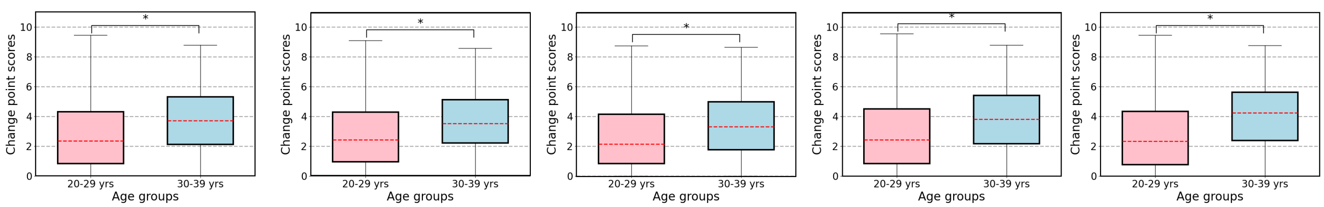
To compare the age-group-wise comparison, we performed the following steps. First, we combined the change point scores obtained from each subject and grouped them as per the age-groups. Next, we checked if the change point scores (for age-groups) follow normal distribution using the Shapiro-Wilk test [34]. We observed that the distribution is not normal (p < 0.05), and the age-groups are not paired; therefore, we performed Mann-Whitney U test [25] to evaluate the difference between age-group-wise change point scores. The comparison for the age-group-wise change point score for all stimuli types are shown in Fig. 4a. The mean change point scores of the two groups (i.e., 20-29 years and 30-39 years) are 2.82 (SD: 2.59) and 3.78 (SD: 2.17), respectively. Also, the medians of change point scores (20-29 years and 30-39 years) are 2.35 and 3.70, respectively. The effect size (Cohen's d) is 0.38. The Mann-Whitney's U test found a significant difference in the age-group-wise change point scores (U = 2247854.0, Z = 0.84, p = 0.00).

We investigate deeper to find out the age-group-wise variation in physiological responses across each type of emotional stimuli



(a) All videos (b) HVHA videos (c) HVLA videos (d) LVHA videos (e) LVLA videos

Figure 3: Evaluation of RQ1 reveals that significant differences exist in the physiological responses (measured in terms of change point score) of male and female participants under continuous emotional stimuli. The significant ($p < 0.05$) differences in change point score across gender are shown in for (a) all stimuli videos, (b) HVHA stimuli videos, (c) HVLA stimuli videos, (d) LVHA stimuli videos, and (e) LVLA stimuli videos.



(a) All videos (b) HVHA videos (c) HVLA videos (d) LVHA videos (e) LVLA videos

Figure 4: Evaluation of RQ1 reveals that significant differences exist in the physiological responses (measured in terms of change point score) of different age-group participants under continuous emotional stimuli. The significant ($p < 0.05$) differences in change point score across age-group are shown in for (a) all stimuli videos, (b) HVHA stimuli videos, (c) HVLA stimuli videos, (d) LVHA stimuli videos, and (e) LVLA stimuli videos.

(i.e., **HVHA, HVLA, LVHA, LVLA**) following the same steps. We show the comparison of the change point score for **HVHA** videos in Fig. 4b. We observed a significant difference in the change point scores across age-group for these videos. The test statistics are as follows (U = 175958.0, Z = 0.82, p = 0.00).

We present the comparison results for other stimuli types (**HVLA, LVHA, LVLA**) in Fig. 4c, 4d, and 4e respectively. We note that for each of the video types (i.e., **HVLA**, **LVHA**, and **LVLA**), there is a significant difference. The test statistics are as follows ({**HVLA**: U = 120753.0, Z = 0.93, p = 0.00}; {**LVHA**: U = 175743.0, Z = 0.84, p = 0.00}; {**LVLA**: U = 97481.0, Z = 0.78, p = 0.00}).

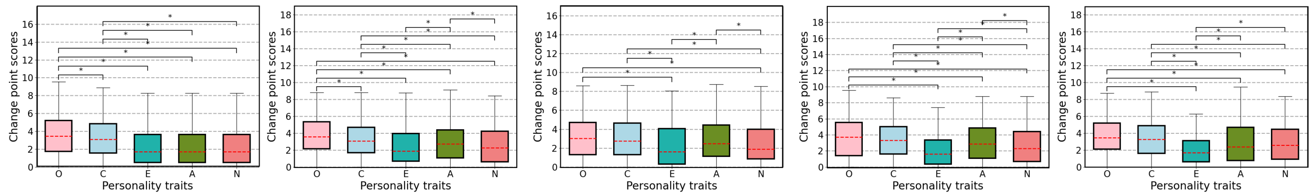
In summary, these findings highlight that physiological responses vary significantly across age-group (older adults (30+ years) have a higher degree of physiological change) for all continuous emotion stimuli. Therefore, a one-size-fits-all approach without considering the age may not be the best choice to detect the emotional variances.

4 RQ2: Influence of Personality on Physiological Response Under Continuous Emotional Stimuli

In this section, we aim to explore how changes in continuous emotional states influence physiological responses among individuals with different personality profiles (**RQ2**). Like previous sections, in this case also, we utilized the change point scores to compare the physiological response variation across the dominant personality traits of the subjects (distribution of the five dominant personalities are presented in Fig. 2b).

To assess the influence of dominant personality on physiological response behavior, we performed the following steps. First, we investigated whether the change-point scores for the five different personality traits followed a normal distribution using the Shapiro-Wilk test [34]. The results indicated that the responses did not adhere to a normal distribution (p < 0.05). Second, given the non-normal distribution and the fact that the five groups (**O**penness, **C**onscientiousness, **E**xtraversion, **A**greeableness, and **N**euroticism) are not paired, we conducted the Kruskal-Wallis test [26] to evaluate the differences in change-point scores across the personality traits. The personality traits which have a significant difference for them the effect size (Cohen's d) varies from 0.11 to 0.56. For example, the highest effect size of **O**penness (mean: 3.61, SD: 2.34) and **E**xtraversion (mean: 2.31, SD: 2.31) is 0.56. The Kruskal-Wallis test found (see Fig. 5a) a significant effect of personality traits on change point scores ($\chi^2(4) = 330.64, p < 0.05$) for all emotional stimuli. Finally, a post-hoc test using Mann-Whitney tests with Bonferroni correction [27] showed the significant differences ($p < 0.05$) between six pairs of personality traits.

Next, we performed stimuli-wise (i.e., **HVHA, HVLA, LVHA, LVLA**) analysis (following the same steps) to understand the physiological response variation across personality traits and presented the results in Fig. 5b, 5c, 5d, and 5e respectively. For **HVHA** videos, the Kruskal-Wallis test found (see Fig. 5b) a significant effect of personality traits on change point scores ($\chi^2(4) = 76.32, p < 0.05$). Finally, a post-hoc test using Mann-Whitney tests with Bonferroni correction showed the significant differences ($p < 0.05$) for all but one pair of personality traits. In the same vein, for other stimulus types (**HVLA, LVHA, LVLA**) also, the Kruskal-Wallis test found a significant effect of personality traits on change point scores for each stimulus type. The test statistics for each stimulus types are as follows ({**HVLA:**$\chi^2(4) = 23.28, p < 0.05$}; {**LVHA:**$\chi^2(4) = 75.81, p < 0.05$}; {**LVLA:**$\chi^2(4) = 60.37, p < 0.05$}). The post-hoc



(a) All videos (b) HVHA videos (c) HVLA videos (d) LVHA videos (e) LVLA videos

Figure 5: Evaluation of RQ2 reveals that significant differences exist in the physiological responses (measured in terms of change point score) of participants with different personality traits under continuous emotional stimuli. The significant ($p < 0.05$) differences in change point score across personality traits are shown in for (a) all stimuli videos, (b) HVHA stimuli videos, (c) HVLA stimuli videos, (d) LVHA stimuli videos, and (e) LVLA stimuli videos. Personality traits Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism are denoted by O, C, E, A, N, respectively.

test using Mann-Whitney tests with Bonferroni correction identifies six, nine, and eight personality-pairs to have a significant difference for **HVLA, LVHA** and **LVLA** stimulus type, respectively. In general, it is observed that persons with dominant personality as **O**penness or **C**onscientiousness are tend to have a higher degree of change in the physiological response in comparison to other dominant personalities.

These findings highlight a significant difference in physiological response behavior among subjects with different personality traits for all types of emotional stimuli videos. As a result, a personality based stratification of users may be needed to obtain superior performance while detecting the high emotional variation moments.

5 Discussion and Future Research Agenda

This early work demonstrated that physiological responses vary significantly across genders, age groups, and personalities. These findings open future research possibilities to optimize the continuous annotation burden in several directions. **First**, the opportunistic annotation can be made more efficient by incorporating user profile details, which are missing in the current state-of-the-art [5]. Notably, in the current user study, we experimented with limited demographic factors (e.g., age: 20 to 39 years only) and personality traits (Big Five). However, by broadening the demographics (e.g., including persons from older age groups) and factoring in other aspects (e.g., education level and cultural differences), the findings of the study can be generalized. **Second**, as noted in the current study, significant changes can be observed across emotional stimuli by considering only HR and GSR. Notably, as these modalities can be collected from off-the-shelf wearable devices, these can be integrated with mobile devices (e.g., smartphones, tablets), which are more widely used for video consumption in daily routines. Therefore, an integrated setup (of wearable and mobile devices) can make the opportunistic annotation in-situ and ubiquitous (not restricted to a static environment). **Third**, in the current analysis, we adopted the same pre-processing strategy for every participant. However, recent literature has highlighted that a personalized approach for different data processing tasks (e.g., winsorization, filtering, segmentation) leads to superior performance when developing data-driven systems using sensor signals [18]. Therefore, by adopting a more nuanced approach for data pre-processing, opportunistic annotation can be made more efficient. **Fourth**, while an opportunistic annotation is a viable option to reduce the continuous annotation effort, there are several crucial unexplored research questions regarding the intervention strategy (for the self-report collection). For example, what would be the degree of interruption if a user is probed for self-reporting during an engaging movie scene? Similarly, what would be an ideal way to collect the annotation (e.g., on-screen by pausing the video or outside screen by providing a nudge on the wearable)? It will also be interesting to investigate the impact of resumption lag (on emotion) once the user's attention is diverted to annotation collection and the attention is reinstated to the movie post-annotation. **Finally**, one of the practical limitations of implementing an opportunistic annotation could be to detect the opportune moments with high precision and recall, and that too in real-time so that the momentary emotional fluctuations are captured. To address this, we aim to develop efficient machine-learning models in our future work.

6 Conclusion

This paper investigates the impact of demographics and personality traits on physiological responses under the influence of continuous emotional stimuli to reduce the continuous emotion annotation effort in video applications. As continuous annotation is burdensome, the continuous annotation effort can be reduced by probing opportunistically at the moments of high physiological signal variation (that are proxies of emotional variation). We conducted a real-world user study (N=31) to validate this by collecting participants' HR, GSR, and continuous emotion annotations (valence, arousal) during video consumption. The analysis of this dataset reveals that physiological responses vary significantly across age, gender, and personality traits. Therefore, the important takeaway from this exploration is that user profiles play an essential role in quantifying the emotional variance manifested in the physiological signals. Thus, there remains scope to optimize the continuous annotation process significantly by considering user demographics and personality traits.

Acknowledgments

This research has been supported by the SURE grant (SUR/2022/001965) of SERB (Science and Engineering Research Board) of the Department of Science & Technology (DST), Government of India, and the CDRF grant (C1/23/152) of BITS Pilani Goa. Snehanshu Saha would like to thank the Anuradha and Prashanth Palakurthi Center for Artificial Intelligence Research (APPCAIR) and ANRF CRG-DST (CRG/2023/003210) for supporting this research.

References

[1] Akhilesh Adithya, Snigdha Tiwari, Sougata Sen, Sandip Chakraborty, and Surjya Ghosh. 2022. OCEAN: Towards Developing an Opportunistic Continuous Emotion Annotation Framework. In *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, IEEE, Pisa, Italy, 9–12.

[2] Samaneh Aminikhanghahi and Diane J Cook. 2017. A survey of methods for time series change point detection. *Knowledge and information systems* 51, 2 (2017), 339–367.

[3] MM Baltes, LL Carstensen, UM Staudinger, and U Lindenberger. 2003. Understanding human development: Dialogues with lifespan psychology.

[4] Paul B Baltes. 1997. On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. *American psychologist* 52, 4 (1997), 366.

[5] Swarnali Banik, Sougata Sen, Snehanshu Saha, and Surjya Ghosh. 2024. Towards Reducing Continuous Emotion Annotation Effort During Video Consumption: A Physiological Response Profiling Approach. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 91 (sep 2024), 32 pages. https://doi.org/10.1145/3678569

[6] Allan Berrocal and Katarzyna Wac. 2018. Peer-vasive computing: Leveraging peers to enhance the accuracy of self-reports in mobile human studies. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. Association for Computing Machinery (ACM), Singapore, 600–605.

[7] John Brebner. 1990. Personality factors in stress and anxiety. *Cross-cultural anxiety* 4 (1990), 11–19.

[8] Cheryl L Carmichael, Harry T Reis, and Paul R Duberstein. 2015. In your 20s it's quantity, in your 30s it's quality: The prognostic value of social activity across 30 years of adulthood. *Psychology and aging* 30, 1 (2015), 95.

[9] Yu-Lin Chang, Yung-Ju Chang, and Chih-Ya Shen. 2019. She is in a bad mood now: leveraging peers to increase data quantity via a chatbot-based ESM. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery (ACM), Taipei, Taiwan, 1–6.

[10] Tara M Chaplin. 2015. Gender and emotion expression: A developmental contextual perspective. *Emotion Review* 7, 1 (2015), 14–21.

[11] Emma Childs, Tara L White, and Harriet de Wit. 2014. Personality traits modulate emotional and physiological responses to stress. *Behavioural pharmacology* 25, 5 and 6 (2014), 493–502.

[12] Lee Anna Clark, David Watson, and Susan Mineka. 1994. Temperament, personality, and the mood and anxiety disorders. *Journal of abnormal psychology* 103, 1 (1994), 103.

[13] Tamlin S Conner, Howard Tennen, William Fleeson, and Lisa Feldman Barrett. 2009. Experience sampling methods: A modern idiographic approach to personality research. *Social and personality psychology compass* 3, 3 (2009), 292–313.

[14] Roddy Cowie, Ellen Douglas-Cowie, Susie Savvidou, Edelle McMahon, Martin Sawey, and Marc Schröder. 2000. 'FEELTRACE': An instrument for recording perceived emotion in real time. In *ITRW Speech-Emotion*. ISCA (International Speech Communication Association), Newcastle, Northern Ireland, UK, 1–6.

[15] Chris Creed, Chris P Bowers, Robert J Hendley, and Russell Beale. 2010. User perception of interruptions in multimedia annotation tasks. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. Association for Computing Machinery (ACM), Reykjavík, Iceland, 619–622.

[16] Surjya Ghosh, Salma Mandi, Sougata Sen, Bivas Mitra, and Pradipta De. 2024. Towards estimating missing emotion self-reports leveraging user similarity: a multi-task learning approach. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM, USA, 1–19.

[17] Jeffrey M Girard and Aidan GC Wright. 2018. DARMA: Software for dual axis rating and media annotation. *Behavior research methods* 50, 3 (2018), 902–909.

[18] Yunjo Han, Panyu Zhang, Minseo Park, and Uichin Lee. 2024. Systematic Evaluation of Personalized Deep Learning Models for Affect Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 4 (2024), 1–35.

[19] Muhammad Asif Hasan, Nurul Fazmidar Mohd Noor, Siti Soraya Binti Abdul Rahman, and Mohammad Mustaneer Rahman. 2020. The transition from intelligent to affective tutoring system: a review and open issues. *IEEE Access* 8 (2020), 204612–204638.

[20] Joel M Hektner, Jennifer A Schmidt, and Mihaly Csikszentmihalyi. 2007. *Experience sampling method: Measuring the quality of everyday life*. SAGE Publications, Inc., Thousand Oaks, CA. https://doi.org/10.4135/9781412984201

[21] Md Adnanul Islam, Md Saddam Hossain Mukta, Patrick Olivier, and Md Mahbubur Rahman. 2022. Comprehensive guidelines for emotion annotation. In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery (ACM), Faro, Portugal, 1–8.

[22] Oliver P John, Sanjay Srivastava, et al. 1999. The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of personality: Theory and research*. Guilford Press, New York, NY, USA, 1–37.

[23] Yubo Kou and Xinning Gui. 2020. Emotion regulation in esports gaming: A qualitative study of league of legends. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–25.

[24] Song Liu, Makoto Yamada, Nigel Collier, and Masashi Sugiyama. 2013. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks* 43 (2013), 72–83.

[25] Henry B Mann and Donald R Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* 18, 1 (1947), 50–60.

[26] Patrick E McKnight and Julius Najab. 2010. Kruskal-Wallis Test. In *The Corsini Encyclopedia of Psychology*. Wiley Online Library, Hoboken, NJ, USA, 1–1.

[27] Patrick E McKnight and Julius Najab. 2010. Mann-Whitney U Test. In *The Corsini Encyclopedia of Psychology*. Wiley Online Library, Hoboken, NJ, USA, 1.

[28] Lakmal Meegahapola, William Droz, Peter Kun, Amalia De Götzen, Chaitanya Nutakki, Shyam Diwakar, Salvador Ruiz Correa, Donglei Song, Hao Xu, Miriam Bidoglia, et al. 2023. Generalization and personalization of mobile sensing-based mood inference models: an analysis of college students in eight countries. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 6, 4 (2023), 1–32.

[29] Varun Mishra, Gunnar Pope, Sarah Lord, Stephanie Lewia, Byron Lowens, Kelly Caine, Sougata Sen, Ryan Halter, and David Kotz. 2020. Continuous detection of physiological stress with commodity hardware. *ACM transactions on computing for healthcare* 1, 2 (2020), 1–30.

[30] Prasanth Murali, Javier Hernandez, Daniel McDuff, Kael Rowan, Jina Suh, and Mary Czerwinski. 2021. Affectivespotlight: Facilitating the communication of affective responses from audience members during online presentations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama, Japan, 1–13.

[31] Cheul Young Park, Narae Cha, Soowon Kang, Auk Kim, Ahsan Habib Khandoker, Leontios Hadjileontiadis, Alice Oh, Yong Jeong, and Uichin Lee. 2020. K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Scientific Data* 7, 1 (2020), 1–16.

[32] Aditya Ponnada, Caitlin Haynes, Dharam Maniar, Justin Manjourides, and Stephen Intille. 2017. Microinteraction ecological momentary assessment response rates: Effect of microinteractions or the smartwatch? *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–16.

[33] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[34] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3/4 (1965), 591–611.

[35] Karan Sharma, Claudio Castellini, Freek Stulp, and Egon L Van den Broek. 2017. Continuous, real-time emotion annotation: A novel joystick-based analysis framework. *IEEE Transactions on Affective Computing* 11, 1 (2017), 1–7.

[36] Tianyi Zhang, Abdallah El Ali, Chen Wang, Alan Hanjalic, and Pablo Cesar. 2020. RCEA: Real-time, continuous emotion annotation for collecting precise mobile video ground truth labels. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, USA, 1–15.