Annotating Affect in the Field: A Case Study on the Usability of a Minimalist Smartwatch User Interface for Affect Annotation

TIM SCHNEEGANS, ANDRY NICLAS, KIRSTEN GREINER, IVO BENKE, ALEXANDER MAEDCHE, and MICHAEL BEIGL, Karlsruhe Institute of Technology

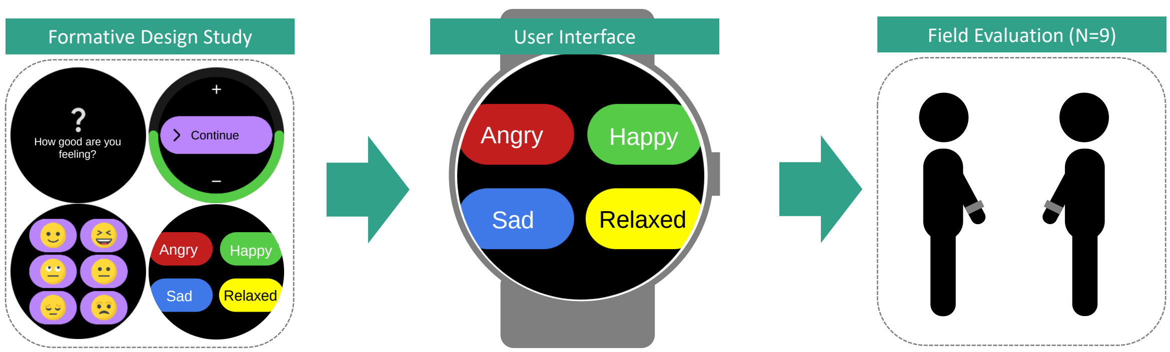

Fig. 1. Process of the design and evaluation steps with the goal of enrolling a Smartwatch UI for affect annotation in the field.

Successful empathetic interaction requires an accurate understanding of the interaction partner's affect dynamics. Self-reported annotations provide a way to better understand affect and empathy in real-life; however, the necessary user interactions for collecting such data must be designed to be as unobtrusive as possible. To address this challenge, we explore the potential of a smartwatch annotation application for affect that aims to minimize user interaction effort while maximizing usability. In a field study conducted as part of a student career fair (N=9), we evaluated the feasibility and usability of our app. Participants reported high usability scores and our data collection successfully captured self-reported affect labels at a high temporal resolution. Our work contributes to the challenge of providing minimal obtrusive applications for the collection of self-reported labels of affective states.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing design and evaluation methods**; **Empirical studies in interaction design**.

Additional Key Words and Phrases: Wearables, User Experience, User Interfaces, Smartwatches, Ambulatory Assessment, Ecological Momentary Assessment, Affect Assessment, Human-Computer-Interaction, Field evaluation

ACM Reference Format:
Tim Schneegans, Andry Niclas, Kirsten Greiner, Ivo Benke, Alexander Maedche, and Michael Beigl. 2018. Annotating Affect in the Field: A Case Study on the Usability of a Minimalist Smartwatch User Interface for Affect Annotation. In . ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Successful empathetic interaction requires an adequate affective response to one's interaction partner, which in turn, requires a precise understanding of their emotions and mood [15, 16]. Affect encompasses both physiological and psychological responses to one's environment [7, 19], which can be classified along the dimensions of arousal and valence [30, 36, 37]. For the purpose of this study, we use the term affect as a general category that includes emotions and mood [37], with emphasis on the psychological component.

Empathetic Human-Computer-Interaction (HCI) can be made possible through the use of affective computing systems [32], as demonstrated by health applications designed to support emotion regulation [9, 17] or emotion-aware conversational agents [8, 24]). Many of these applications rely on subjective data annotations, where the user manually responds to a query about their current affect state. However, daily duties and distractions often hinder the motivation of those users to provide such data with a high degree of frequency and accuracy [18]. Minimizing the necessary user interaction would not only increase user comfort but also enable researchers to collect annotations on a higher temporal resolution [14, 34].

Smartwatches are an attractive tool for collecting annotation data in the field. Their quickly accessible display [5] demands minimal distraction from everyday life. A large user base (over 127.5 Million sold units in 2021), suggests a widespread user acceptance [20, 39]. Furthermore, smartwatches can leverage integrated sensors for light, heart rate, or acceleration. Yet to fully untap this potential, we miss crucial knowledge about best practices for UI design and data collection of affective experiences.

The contribution of this paper is the evaluation of a smartwatch application for affect annotation and data collection in the field. In section 3, we detail our iterative design process with draws on related work and pretests. In section 4, we present a small-scale field evaluation during a student career fair. Career fairs provide a relevant use case for affect annotation because they contain intense social interactions and other challenging distractions for usability evaluations and diverse affect states (e.g., excitement, enthusiasm) that impact job seekers' success [10, 13, 25]. The results demonstrate the ease of use and the potential of this prototype for collecting affect with high temporal density. In section 5, we explore the potential implications of our findings for applications beyond career fairs, suggest ways to enhance future UI studies, and propose further UI ideas minimal obtrusive data annotation tools.

2 RELATED WORK

The quality of manually reported data is heavily dependent on the user interface of the data collection device [1]. Traditional affect assessment tools present users with verbal self-descriptions (e.g., *'I feel tense'*) and require them to rate their affective state on a numeric scale [31, 40]. Non-verbal items, like emojis [22] or manikins [11, 19], were proposed as a simpler alternative. There are also attempts to assess arousal and valence with one user interaction, where the user navigates to a specific arousal-valence combination on a two-dimensional grid [3, 14].

Ponnada et al. [34] raised the idea of using smartwatches for minimizing user interaction for affect annotation. The authors demonstrated that smartwatches can achieve the same validity and response rates as established smartphone based studies. Furthermore, existing research has proposed specific UIs for smartwatch-based affect assessment, via numeric items, emojis, [6], and text-based items [34].

3 DESIGN PROCESS AND DEVELOPMENT

Figure 1 provides a preview of our iterative design process. We first focused on eliciting qualitative data for the purpose of identifying specific design and interaction elements that are appropriate for affect assessment. Annotating data using smartwatches is challenging because of the limited touchscreen space, the limited patience of the user, as well as distractions in everyday life. This brings the necessity to cut down the amount of collectable data while maintaining validity and providing input items for the full range of affect states.

3.1 Formative Design Study

In a formative pretest, we compared seven prototypes in regards to the usability and validity for measuring affect. From qualitative user feedback, we learned that word descriptions are less ambiguous than emojis [6, 22, 26] and that a multiple choice option with short words allows for quicker user interaction compared to numerical values for each affect dimension. Based on these findings, we developed a final UI design which was implemented and evaluated at a career fair. The application was implemented for Android WearOS, using the Compose toolkit for WearOS. The code was written in Kotlin for the Samsung Galaxy Watch 5 Pro. To ensure data privacy, we used a local database (Room library) and only used WiFi for downloading the data via the Android Debug Bridge after data collection has concluded.

3.2 The Smartwatch User Interface

Figure 2a depicts a visualization of the final UI. To reduce the interaction effort, we enabled the user to simultaneously assess the arousal and valence dimension by choosing an adjective that is mapped along both dimensions to the x- and y-axis on the screen [3, 14]. Figure 2b illustrates the mapping of each item to the position on the arousal-valence grid as proposed by [33] and applied by [2, 23]. In order to facilitate a more granular input for describing affect, we combined two interaction steps in an adaptive decision-tree-like fashion. After responding to the prompt, the user defines the rough quadrants of the affect-valence-grid by choosing among the four items the one that most closely matches their affective state (step 1). Subsequently (step 2), the user chooses among four more items to further specify the affect state. For example, in step 1, the user indicates that they feel "happy", which means high arousal, positive valence. In step 2, the user specifies the affect state ("happy", "excited", "aroused", "delighted"). These items all describe high arousal and positive valence but differ in their exact degree of these dimensions. We decided to offer no neutral state in order to avoid biases towards neutral options [29].

4 FIELD EVALUATION

4.1 Sample

Data was collected over a period of four-hours on a student career fair, in sync with the start and end of the career fair (1.30 pm to 5.30 pm; Table 1). N=9 (3 female) students participated voluntarily, receiving a fixed compensation of 15 € and a variable compensation of 50 cents for each prompt response. Originally, ten participants were recruited out of approx. 80 career fair visitors but one was removed due to technical problems. Most participants were 18-25 years old (n=8); one was 25-30. This person was the only one who reported wearing a smartwatch regularly (24 hours a day) (user ID 8, Table 1).

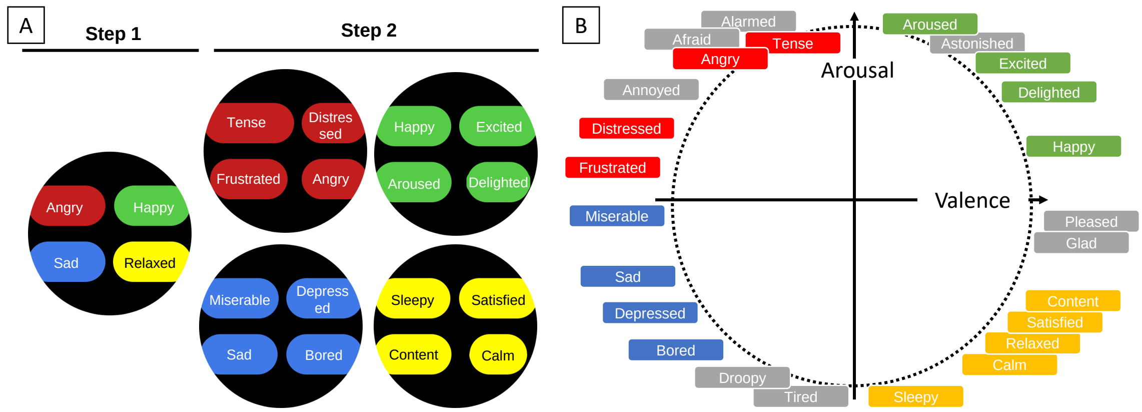

Fig. 2. (A) Screenshots from the UI for each interaction. In step 2, only one screen is shown to the user depending on the chosen item in step 1. The original German items were translated to English for this illustration. (B) Alignment of the items along the arousal-valence dimensions as proposed by [33].

4.2 Measures

4.2.1 Usability. At the end of the career fair, we measured usability with the System Usability Scale (SUS) [28] and the Post-Study System Usability Questionnaire (PSSUQ) [27].

4.2.2 Affect Assessment via Smartwatch. In regular intervals, our users were prompted to indicate their current affect state via the UI as proposed above. In a prompt, the smartwatch silently vibrates and shows the message "Time to interact". The users were instructed to tap on the notification to see the actual annotation screen (Figure 2). Although the prompts were programmed in 15 min intervals, internal task prioritization mechanisms of the WearOS system delayed them to 18.9 min on average (SD=1.3 min). Our users received on average M=10.8 notifications (SD=2.5) and responded to almost all of them (M=9.6, SD=2.3) (Table 1). Five participants missed one prompt, three participants missed two prompts, and one missed no prompt. In summary, we collected data on the following events with the respective timestamps: a prompt was sent, the user taps on a prompt, the user inputs their affective state, and pause and resume actions made by the user.

4.2.3 Affect Assessment via Surveys. At the end of the career fair, we measured affect with the German version of the PANAS [12, 40] questionnaire, which serves as a gold standard comparison for the smartwatch assessment.

4.3 Procedure

The career fair involved two steps: first, the students had to present the results of a semester software development team project in a short pitch-session to a wide audience of peers and possible future employers. Later, at around 2:30 pm, they engaged in free networking opportunities at fair stands of the resulting software prototypes as well as partner companies. Students may encounter diverse affective states in this environment, including stressful (e.g., while presenting) and positive experiences (e.g., finding interesting job opportunities). Furthermore, the event provides numerous social interactions and other distractions, making it a challenging setting for a usability evaluation.

Table 1. Affect Annotations aligned by time, average PANAS scores for positive (PA) and negative affect (NA), and number of responded notifications (N) for each participant (ID). Colors indicate arousal-valence combinations (green: positive valence, high arousal; yellow=positive valence, low arousal; blue: negative valence, low arousal; red: negative valence, high arousal. x=annotation prompts that were not answered. Con=Content, Aro=Aroused, Hap=Happy, Sat=Satisfied, Cal=Calm, Ang=Angry, Bor=Bored, Exc=Excited.

	Smartwatch Annotations																	PANAS	
ID	1.30 pm		2.00 pm		2.30 pm		3.00 pm		3.30 pm		4.00 pm		4.30 pm		5.00 pm		N	NA	PA
1								Con	Aro	Aro	x	Hap	Hap		Con	Sat	7	1.6	4.1
2		Cal		Aro	Ang	Sat	Con	Hap		Con	Sle	Con	Sat		Con		11	2.3	2.5
3		Sle	Sle	Aro		Cal	Aro		Hap	Hap	Exc	Exc	Con		Hap	Exc	13	2.0	3.8
4	Sat	Sat		Cal	Con	Sat	Sle	x	Con		Sle	Con	Exc		Con		10	1.0	2.9
5					x	Bor	Bor	Aro	Sat	Con	Con	Con	Con	Exc		Exc	10	1.2	4.2
6	Ang		Ang	Con	Hap	Sle		Cal	x	Sat	Cal		Cal	Hap	Con	Bor	12	1.3	3.3
7								Sat	Con	Con	Con		Con	x		Cal	6	1.2	3.8
8										Exc	Cal	Exc	Cal	Exc	Exc	Dep	7	1.9	3.4
9				Sat		Cal	Aro	Con	Sat	Con		Con	Con	Con		Hap	10	1.4	3.7

4.4 Results

4.4.1 Usability. The participants indicated a very high usability for the SUS (M=86.7, SD=8) and PSSUQ (M=1.7, SD=0.4). The PSSUQ subscales showed the highest values for usefulness (M=1.3, SD=0.32) and the lowest values for information quality M=2.5 (SD=1.18) (interface quality: M=1.7, SD=0.75).

4.4.2 User Interaction and Affect Assessment via Smartwatch. The average response time for a prompt was M=7.6 s (SD=4.5 s). Table 1 visualizes affect measures in the form of the smartwatch annotation across time as well as the PANAS responses for each participant. Five participants missed one notification. Because the annotation intervals were often longer than 15 min (subsection 4.3), intervals that cannot be temporally assigned to any annotation prompt are marked as white space in the table. Almost all participants selected positive valence - high arousal items (green boxes), such as "happy" and "relaxed". Moreover, some participants started the data collection later at the career fair (with IDs 1, 7 and 8). On average, the time window for receiving annotation prompts lasted for M=174 min (SD=54 min). Moreover, the users walked on average M=2350 steps (SD=1542) during the career fair, which indicates strong movement activity.

4.4.3 Affect Assessment via Survey. The results from the PANAS questionnaire indicate low negative affect (M=1.5, SD=0.9) and high positive affect (M=3.5, SD=1.0) (Table 1). This aligns with the mostly positive valence affect annotations of the smartwatch measure. However, we also find differences between the final questionnaire results at the end of the career fair and the affect dynamics throughout the career fair. For example, two participants (with IDs 6 and 8) occasionally indicated negative valence annotations (participant 6: "bored"; participant 8: "depressed") during the career fair but positive affect scores in the PANAS survey.

5 DISCUSSION

In this work, we evaluated a UI for affect assessment on a smartwatch in order to collect affect annotations at a high temporal frequency while creating minimal distractions for the user. With our results, we provide design knowledge for non-invasive and easy-to-use data collection on affect states and emotions in the field. Our prototype received high usability scores, despite the fact that our participants were distracted by the hustle and bustle of the career fair and were walking around a lot, as indicated by the step counter of our smartwatches. Moreover, our data suggest an overall alignment between the number of positive affect annotations and the positive affect scores in the survey.

Furthermore, we manage to go beyond the one-time assessment of the survey and zoom into the temporal dynamics of the specific affect annotations on the smartwatch on a student career fair. For example, from our collected data, we can interpret individual affect changes: At the beginning of the career fair, participants engaged in the pitch session, which could explain the occurrence of "tense" annotations at the beginning during their presenter role and subsequent "sleepy", "calm", or "satisfied" annotations after returning to the audience.

Our method has broader applications beyond career fairs, as it can be utilized in designing affective computing interfaces and emotion-adaptive systems [3, 24, 35]. Additionally, it has potential use cases in assessing affect dynamics in daily workplace routines [41] and in the daily lives of patients with psychiatric conditions [38].

The adaptive structure items in a decision-tree-like fashion of our UI facilitates a larger and therefore more granular set of emotion labels compared to other design proposals [3, 6, 14]. Thereby, our UI maintains a high usability. In the future, we want to further enhance the adaptiveness of the UI by personalizing the items to the user and the context. For example, a well-trained user does not need to know the meaning of each item at full length but could use symbolic values instead. Furthermore, by leveraging prior user annotations or applying affective computing techniques on cardiovascular data, only choice options that the user is highly likely to select can be suggested.

5.1 Limitations and Future Research

To improve the reliability of our assessment approach, the mapping of single-word descriptions to self-reported numerical arousal-valence measures should be further refined on a larger and up-to-date dataset. Although, we chose our items based on solid psychological literature [33], the everyday perception could have changed since the 1990s and also depends on the user group.

Furthermore, we could not exactly identify the reason for missed annotation prompts. It is possible that the users did not perceive the vibration of the smartwatch or that they noticed but ignored the prompts. The high response rate suggests this as a marginal problem for this study. However, future research should further investigate the user acceptance and response rates in diverse everyday situations.

Future research should also explore the impact of design elements such as color or the organization of items on user behavior and harmonize it with the user's intuitive understanding. For example, western readers have a visual bias towards the upper left options of a screen [4]. Another optimization option would be to align the color scheme with the user's intuitive understanding (e.g., using red for negative valence items [21]). In this study, we opted for selecting colors for the choice options with the focus of maximizing their visible discernibility.

Our findings on the response rates of our application should be considered under the limitation that we provided financial incentives as an extrinsic motivator for each user interaction. To enhance intrinsic motivation for user interaction, our interface could be combined with user feedback that provide subjective benefits to the user, for example in the form of a self-monitoring system on emotions.

6 SUMMARY

To investigate emotional dynamics and emotion-related skills, such as empathy, it is important to understand affective states during real-life activities. For this, it is necessary to collect self-reported affect annotations. In this paper, we tested an easy-to-use smartwatch application for affect annotations in a field study. The results confirm the usability and feasibility of collecting affect information at a high frequency in real-life events. In the future, we plan to explore data collection procedures that improve quality by focusing usability and user motivation, temporal resolution by minimizing the interaction time, and ecological validity by minimizing distractions from everyday life.

REFERENCES

[1] Janko Ahlbrandt, Michael Henrich, Bernd A Hartmann, Bettina B Bundschuh, Julia Schwarz, Joachim Klasen, and Rainer Röhrig. 2012. Small cause–big effect: improvement in interface design results in improved data quality–a multicenter crossover study. In *Quality of Life through Quality of Information*. IOS Press, 393–397.

[2] Omid Alemi, William Li, and Philippe Pasquier. 2015. Affect-expressive movement generation with factored conditional Restricted Boltzmann Machines. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. 442–448. https://doi.org/10.1109/ACII.2015.7344608

[3] Sebastian Alpers and Kuderna-Iulian Benţa. 2021. SenseGraph: Affect Self-monitoring and Tagging Tool with Wearable Devices. In *2021 International Conference on Computer Communications and Networks (ICCCN)*. IEEE, 1–6.

[4] Dianna R Amasino, Nicolette J Sullivan, Rachel E Kranton, and Scott A Huettel. 2019. Amount and time exert independent influences on intertemporal choice. *Nature human behaviour* 3, 4 (2019), 383–392.

[5] Daniel L Ashbrook, James R Clawson, Kent Lyons, Thad E Starner, and Nirmal Patel. 2008. Quickdraw: the impact of mobility and on-body placement on device access time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 219–222.

[6] Anja Bachmann, Christoph Klebsattel, Andrea Schankin, Till Riedel, Michael Beigl, Markus Reichert, Philip Santangelo, and Ulrich Ebner-Priemer. 2015. Leveraging smartwatches for unobtrusive mobile ambulatory mood assessment. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 1057–1062.

[7] Lisa Feldman Barrett and Eliza Bliss-Moreau. 2009. Affect as a psychological primitive. *Advances in experimental social psychology* 41 (2009), 167–218.

[8] Ivo Benke, Michael Thomas Knierim, and Alexander Maedche. 2020. Chatbot-Based Emotion Management for Distributed Teams: A Participatory Design Study. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 118 (oct 2020), 30 pages. https://doi.org/10.1145/3415189

[9] Ivo Benke, Maren Schneider, Xuanhui Liu, and Alexander Maedche. 2022. TeamSpiritous - A Retrospective Emotional Competence Development System for Video-Meetings. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 392 (nov 2022), 28 pages. https://doi.org/10.1145/3555117

[10] Silvia Bonaccio, Natalie Gauvin, and Charlie L Reeve. 2014. The experience of emotions during the job search and choice process among novice job seekers. *Journal of Career Development* 41, 3 (2014), 237–257.

[11] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.

[12] Bianka Breyer and Matthias Bluemke. 2016. Deutsche version der positive and negative affect schedule PANAS (GESIS panel). (2016).

[13] Stéphane Côté, Alan M Saks, and Jelena Zikic. 2006. Trait affect and job search outcomes. *Journal of Vocational Behavior* 68, 2 (2006), 233–252.

[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 *ISCA tutorial and research workshop (ITRW) on speech and emotion*.

[15] Benjamin MP Cuff, Sarah J Brown, Laura Taylor, and Douglas J Howat. 2016. Empathy: A review of the concept. *Emotion review* 8, 2 (2016), 144–153.

[16] Carol M Davis. 1990. What is empathy, and can empathy be taught? *Physical therapy* 70, 11 (1990), 707–711.

[17] Vipula Dissanayake, Vanessa Tang, Don Samitha Elvitigala, Elliott Wen, Michelle Wu, and Suranga Nanayakkara. 2022. Troi: Towards Understanding Users Perspectives to Mobile Automatic Emotion Recognition System in Their Natural Setting. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 203 (sep 2022), 22 pages. https://doi.org/10.1145/3546738

[18] Ulrich W Ebner-Priemer and Günther Sawitzki. 2007. Ambulatory assessment of affective instability in borderline personality disorder: The effect of the sampling frequency. *European Journal of Psychological Assessment* 23, 4 (2007), 238.

[19] Maria Egger, Matthias Ley, and Sten Hanke. 2019. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science* 343 (2019), 35–55.

[20] Gartner Inc. 2021. Gartner Forecasts Global Spending on Wearable Devices to Total $81.5 Billion in 2021. https://www.gartner.com/en/newsroom/press-releases/2021-01-11-gartner-forecasts-global-spending-on-wearable-devices-to-total-81-5-billion-in-2021.

[21] Avery N Gilbert, Alan J Fridlund, and Laurie A Lucchina. 2016. The color of emotion: A metric for implicit color associations. *Food Quality and Preference* 52 (2016), 203–210.

[22] Sara R Jaeger, David Jin, Grace S Ryan, and Joachim J Schouteten. 2021. Emoji for Food and Beverage Research: Pleasure, Arousal and Dominance Meanings and Appropriateness for Use. *Foods* 10, 11 (2021), 2880.

[23] Sara R Jaeger, Christina M Roigard, and Sok L Chheang. 2021. The valence× arousal circumplex-inspired emotion questionnaire (CEQ): Effect of response format and question layout. *Food Quality and Preference* 90 (2021), 104172.

[24] Shin Katayama, Akhil Mathur, Marc Van den Broeck, Tadashi Okoshi, Jin Nakazawa, and Fahim Kawsar. 2019. Situation-aware emotion regulation of conversational agents with kinetic earables. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 725–731.

[25] Eunjin Kim and Bora Lee. 2022. The role of emotion in job search behavior among college students. *Journal of Career Development* 49, 1 (2022), 90–106.

[26] Gaku Kutsuzawa, Hiroyuki Umemura, Koichiro Eto, and Yoshiyuki Kobayashi. 2022. Classification of 74 facial emoji's emotional states on the valence-arousal axes. *Scientific Reports* 12, 1 (2022), 398.

[27] James R Lewis. 1992. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the human factors society annual meeting*, Vol. 36. Sage Publications Sage CA: Los Angeles, CA, 1259–1260.

[28] James R Lewis. 2018. The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction* 34, 7 (2018), 577–590.

[29] Petra Lietz. 2010. Research into questionnaire design: A summary of the literature. *International journal of market research* 52, 2 (2010), 249–272.

[30] Albert Mehrabian. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology* 14, 4 (1996), 261–292.

[31] Albert Mehrabian and James A Russell. 1974. *An approach to environmental psychology.* the MIT Press.

[32] Rosalind W Picard. 2000. *Affective computing.* MIT press.

[33] Robert Ed Plutchik and Hope R Conte. 1997. *Circumplex models of personality and emotions.* American Psychological Association.

[34] 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.

[35] Kiran K Rachuri, Mirco Musolesi, Cecilia Mascolo, Peter J Rentfrow, Chris Longworth, and Andrius Aucinas. 2010. EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. 281–290.

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

[37] James A Russell. 2003. Core affect and the psychological construction of emotion. *Psychological review* 110, 1 (2003), 145.

[38] Timothy J Trull, Sean P Lane, Peter Koval, and Ulrich W Ebner-Priemer. 2015. Affective dynamics in psychopathology. *Emotion Review* 7, 4 (2015), 355–361.

[39] Vantage Market Research. 2021. Smartwatch Market Size USD 80.1 Billion by 2028. https://www.vantagemarketresearch.com/industry-report/smartwatch-market-1507.

[40] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.

[41] Frank H Wilhelm and Paul Grossman. 2010. Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological psychology* 84, 3 (2010), 552–569.

ACKNOWLEDGMENTS

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2739/1 – Project Nr. 447089431 – Research Training Group: KD²School – Designing Adaptive Systems for Economic Decisions