Understanding Implicit User Feedback from Multisensorial and Physiological Data: A case study

Suni-Lopez, Franci; Condori-Fernandez, Nelly; Catala, Alejandro

published in
ICSEW '20
2020

DOI (link to publisher)
10.1145/3387940.3391466

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Suni-Lopez, F., Condori-Fernandez, N., & Catala, A. (2020). Understanding Implicit User Feedback from Multisensorial and Physiological Data: A case study. In *ICSEW '20: Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops* (pp. 563-569). Association for Computing Machinery, Inc. https://doi.org/10.1145/3387940.3391466

Understanding Implicit User Feedback from Multisensorial and Physiological Data: A case study

Franci Suni-Lopez
Universidad Nacional de San Agustín de Arequipa, Perú
fsunilo@unsa.edu.pe

Nelly Condori-Fernandez
Universidade da Coruña
Vrije Universiteit Amsterdam
Coruña, Spain, Amsterdam, The Netherlands
n.condori.fernandez@udc.es
n.condori-fernandez@vu.nl

Alejandro Catala
Centro Singular de Investigacion en Tecnoloxias Intelixentes, Universidade de Santiago de Compostela
Spain
alejandro.catala@usc.es

ABSTRACT

Ensuring the quality of user experience is very important for increasing the acceptance likelihood of software applications, which can be affected by several contextual factors that continuously change over time (*e.g.,* emotional state of end-user). Due to these changes in the context, software continually needs to adapt for delivering software services that can satisfy user needs. However, to achieve this adaptation, it is important to gather and understand the user feedback. In this paper, we mainly investigate whether physiological data can be considered and used as a form of implicit user feedback. To this end, we conducted a case study involving a tourist traveling abroad, who used a wearable device for monitoring his physiological data, and a smartphone with a mobile app for reminding him to take his medication on time during four days. Through the case study, we were able to identify some factors and activities as emotional triggers, which were used for understanding the user context. Our results highlight the importance of having a context analyzer, which can help the system to determine whether the detected stress could be considered as actionable and consequently as implicit user feedback.

KEYWORDS

Actionable emotion, Implicit user feedback, Physiological data, Context information, Case study

ACM Reference Format:
Franci Suni-Lopez, Nelly Condori-Fernandez, and Alejandro Catala. 2020. Understanding Implicit User Feedback from Multisensorial and Physiological Data: A case study. In *IEEE/ACM 42nd International Conference on Software Engineering Workshops (ICSEW'20), May 23–29, 2020, Seoul, Republic of Korea.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3387940.3391466

1 INTRODUCTION

Nowadays, software systems have become more complex but at the same time adaptable due to the continuous evolution of information technologies such as wearable sensors. However, to achieve the (self-) adaptation of a system, it is required to obtain information about any response of the interaction between the system and the user (*i.e.,* user feedback) [40]. According to Mezhoudi and Vanderdonckt, the usefulness of this information depends on its consistency, credibility, and accuracy [34]. Given user feedback has become a resource of valuable information for software developers, companies or users that are looking for an opinion of the quality of a service or product (*e.g.,* [32]), it has been largely investigated by different computer science fields, especially in Human-Computer Interaction [22, 28, 34, 39, 51] and Software Engineering [16, 33, 35].

Hence, one of the main issues is how to collect useful user feedback which is both accurate and credible [34]. In this direction, different approaches have been investigated (*e.g.,* [13, 18, 26, 36]). For example, some works argue that implicit user feedback —an indirect or inferred reaction of the user upon his experience using a software service (*e.g.,* elapsed time, emotions, mouse movement, mouse clicks and scrolling) [35]—can be used as a suitable source in many cases without distracting users from their primary tasks. Besides, the work in [11] introduced the idea of measuring physiological emotions as an alternative source to empower the adaptability of software services at run-time. As a result of that preliminary research, authors argued that negative emotions (*i.e.,* stress) could also be considered as implicit user feedback for enhancing adaptability and user experience in service-based mobile applications [9].

We developed a real-time stress detector that was initially evaluated within an experimental setting in [31]. Nevertheless emotions are subjective and highly context-dependent [3, 25]. Thus, conducting empirical studies in real-life environments is much more challenging, especially since users are exposed to diverse stimuli under uncontrolled conditions or settings.

In this paper, we present a case study with the aim of investigating whether negative emotions (*i.e.,* stress) can be used as implicit user feedback. Our case focuses on a tourist visiting a city abroad, who interacted with a persuasive mobile app and was exposed to different situations during four days, while context and physiological data were gathered using wearable sensors in combination with other smartphone sensors.

The paper is organized as follows. Section 2 discusses related work on user feedback. The description of the case study is presented in Section 3. The results and design of the context analyzer are discussed in Section 4. Section 5 provides the threats to validity. Finally, conclusions and future works are presented in Section 6.

ACM ISBN 978-1-4503-7963-2/20/05…$15.00
https://doi.org/10.1145/3387940.3391466

2 RELATED WORK ON USER FEEDBACK

The assessment of user feedback is important for self-adaptation and software evolution, which is concerned about involving old designs to generate new ones [16]. In this context, the user feedback is interpreted as relevant information from any response about the interaction between the software system and the user [33]. During many years, user feedback has been mainly used by recommendation systems [35], depending largely on explicit feedback (*e.g.,* filling questionnaires, sending a report about an error, suggestion of new functionalities, or ratings) that users can provide. Nonetheless, users are not interested to give feedback, or they could even give false feedback; for that reason, explicit feedback is sometimes difficult to obtain [20].

On the other hand, implicit feedback (*e.g.,* facial expressions, time on page, click-stream, scrolling, mouse movement tracking) raised as a big data source of information. In this line, researchers have proposed different approaches to deal with implicit feedback. For example, Peska [39] proposed a model of relevant contextual features for collecting implicit user feedback; as a result, a dataset of feedback of real e-commerce users was published. Lee et al. [27] proposed to construct explicit feedback data (pseudo ratings) from implicit feedback data for an e-commerce recommender system. Leiva [28] captured touch events (as implicit feedback) for adapting, rearranging and restyling the interacting items of a web browser. Claypool et al. [8] analyzed the correlation between implicit and explicit feedback, they developed a web browser for collecting implicit feedback(*i.e.,* elapsed time, mouse movement, mouse clicks and scrolling) and explicit feedback (ratings). Additional examples can be found in the existing literature of implicit feedback [20, 21, 45, 48]. Some works also present mixed models combining both implicit and explicit feedback, as in [29], [50], [30], or [23].

In contrast, our proposal recognizes physiological emotions for gathering implicit user feedback. Although, the recognition of human emotions still has many challenges to address. For instance, modeling the variables that intervene and influence the user context is one of these challenges. In this respect, many works have been proposed [14, 15, 17, 19, 36, 42, 46]; Siirtola [44] studied the stress detection as binary classification (stressed vs. non-stressed), based only on Empatica E4 signals, where amusement and relaxed states were combined as one. Schmidt et al.[43] work with 2-classes (stressed vs. non-stressed) and 3-classes (stress, amusement, and relaxed) using the recorded signals from the chest and wristband. Panure and Sonawani [38] proposed a person-specific model to detect stress (using artificial neural network and support vector machine classifiers with LASSO regression models) based in the E4-wristband signals and smartphone. However, most approaches are focused on predefined tasks and not considering other elements in the user context; furthermore, none of them have addressed emotion recognition related to the interaction with a software service (actionable emotion) and they only use stressors in their experiments.

3 CASE STUDY

3.1 Method

The case study aims to investigate whether negative emotions (*i.e.,* physiological stress) can be used as implicit user feedback. From this goal, the following research question (RQ) has been formulated: *Can physiological data be considered and used as a form of user feedback?*

3.2 Analysis unit

The unit of analysis of our case study is a voluntary tourist, who used a smartphone and a wearable wristband device to track physiological data. The subject was selected by convenience. The Persona method has been applied for identifying and capturing significant details of the subject (tourist). These details include personal information, goals, preferences, challenges, and skills. We followed the steps provided by Adlin [1]. Regarding the personal information of the subject, he is a 49-years-old Peruvian math teacher. His native language is Spanish, and he does not speak any other language. As for his personality, the Big-Five Inventory (BFI) questionnaire [24] was administered [1], the subject reported a strong preference to openness (45 points) and a weak inclination to agreeableness (32 points) and neuroticism (27 points). Regarding his technological skills, the subject uses with certain frequency the use of some applications like social networks (*e.g.,* Facebook) and e-mail. Moreover, he does not have any difficulty to use new technological devices since he is always open to learn.

Regarding his preferences, he loves the Roman culture and he also has a particular interest in the Vatican City since he is Catholic. This is why he visited Rome (Italy) by four days in the last days of August. The subject was diagnosed with a *non-allergic rhinitis* that affects the nasal mucus and produces sneezing, itching, obstruction, nasal secretions, and sometimes lack of smell [37]. For that reason, he was prescribed with oral antihistamines twice by day (one pill in the morning and other in the evening).

3.3 Treatment and setting

The subject accepted to use the mobile app, Health Care Reminder [47], for reminding him the intake of his pills. This application was configured based on the prescription given by the doctor. The app implements an intake reminder service by means of two persuasive messages per day: one in the morning (08:00 hours) and the second in the night (20:00 hours).

The case study is conducted in the natural or real-life environment where the subject is exposed to a diverse set of stimuli under uncontrolled conditions and diverse backgrounds, either indoor or outdoor. Moreover, with the purpose of investigating whether the stress can be used as user feedback, the settings of the mobile app (Health Care Reminder) were intentionally set as follows: i) volume, varied between high (*i.e.,* maximum volume) and medium; ii) frequency of the delivery, where a persuasive message is replayed five times, and iii) content of the delivered messages, which are emitted with different persuasiveness levels. The levels correspond to the authority and consensus principles defined by Cialdini [7] (see [47] for details about how the persuasive messages are composed).

Regarding the mobile app, it has been developed in Android and was installed on his mobile phone (a Samsung Galaxy J5). The messages were spoken via voice using the voice assistant for Android

[1] It includes items from three dimensions (*i.e.,* Agreeableness, Neuroticism, Openness), which are related to the use of technology. The Personality test can be found in shorturl.at/bjHQ4

[47]. Figure 1 shows an overview of the setting, where the user uses a wearable device, an E4-wristband, and interacts with the mobile App. Also, different aspects of the context are illustrated. The observer is a researcher (and a close friend of the user as well), who takes notes about the presence of potential emotional triggers and record the information of the activities of the subject.

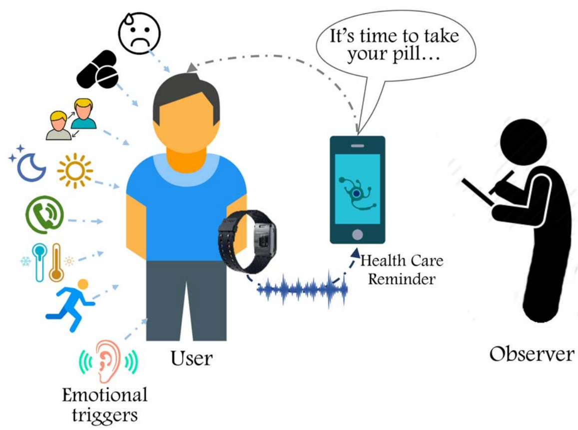

Figure 1: Overview of the case study setting: End-user, Observer, E4-wristband, mobile App, and context.

3.4 Data collection and measurement

The data collection was carried out by means of different instruments/techniques such as:

- the E4-wristband device for collecting i) EDA (Electrodermal Activity) signals, which were then processed by the stress detector implemented in [31]: in short, the detector applies a median filter over a moving window as pre-processing to clean raw signals, and then the algorithm aggregates sampled readings on the filtered input signal in order to carry out a discretization that outputs values between 1 and 5. These values are interpreted as a stress variation from completely relaxed (1) to maximum arousal (5). As a result, data is transformed into time series with discrete sequences. The approach uses a change detection algorithm based on ADaptive WINdowing (ADWIN) method in [4] to eventually assign the stress label. Details on the implementation of the stress detector can be found in [31]. ii) Skin temperature; and iii) accelerometer data used to derive the physical activity changes, according to the method proposed in [2].
- the smartphone's microphone sensor is used to gauge the environmental noise. Captured sound is transformed into a decibels (dB) scale; then thresholding is applied to detect noise [2] ($> 80dB$ [12]).
- A four-day diary study used to take notes about the external context (*e.g.,* weather temperature, locations, performed activities by the subject). This information will allow us to analyze mismatches between the stress detector output and self-reported stress.
- An oral brief inquiry was conducted immediately after each message delivered by the mobile app, by asking to the subject (tourist) about his stress perception. The question was formulated as follows: *Did you feel stress while the reminder was playing?*

The data collection started 15 minutes before the delivery of the first persuasive message, and it finalized five minutes after the second message delivered each day [3]. In the following subsections, we present the activities performed by the subject (Figures 2 and 3 illustrate the activities of day one and four respectively), as well as the stress measurement using the EDA signals that were collected along the four-day study (Table 1 summarize the stress levels experimented by the subject along the four days).

Table 1: Stress levels (SAX representation from 1 to 5) of the subject along the four days.

	SAX representation				
	1 Completely relaxed	**2**	**3**	**4**	**5** Stressed
Day 1	0%	79%	2%	7%	12%
Day 2	34%	14%	18%	9%	25%
Day 3	21%	11%	19%	36%	13%
Day 4	0%	70%	7%	7%	16%

3.4.1 Day one. The collection of data started at noon. Overall, the subject activities were performed during the travel (*e.g.,* traveling by train, sitting on an airplane) from his temporary home (The Netherlands) to Rome (Italy). In this first day, EDA signals were collected three times at different intervals of time. In Figure 2 these intervals are represented with three coloured lines (orange, yellow and green). For instance, the first collection was taken from 12:43 to 13:20 hours, during the travel by train. From Figure 2, we can clearly see that EDA values started to change drastically in the third interval. In this last interval, the mobile app (Health Care Reminder) emitted one message, with an authority level of persuasiveness and a maximum volume. During day one, data was gathered for 2.5h. We found that the subject was stressed about 12% of this total time (see Table 1).

3.4.2 Day two. During the morning of the second day, the subject spent most of his time resting at the hotel. However, during the afternoon, he walked around downtown Rome. The first persuasive message was about consensus level with high volume, and at the moment of delivery the subject was walking in his room. Both the physical activity (walking) and the hot weather (27 °C) influenced on the EDA values. Then, in the afternoon, the subject visited the museum *Sacrario delle Bandiere delle Forze Armate*, where had to pass through a security checkpoint. The second message was of consensus level with medium volume, delivered when the subject was sitting to have dinner at a traditional Rome restaurant. As

[2] http://www.aes.org/

[3] Raw data of the subject activities of each day can be found at https://goo.gl/hGwPpk

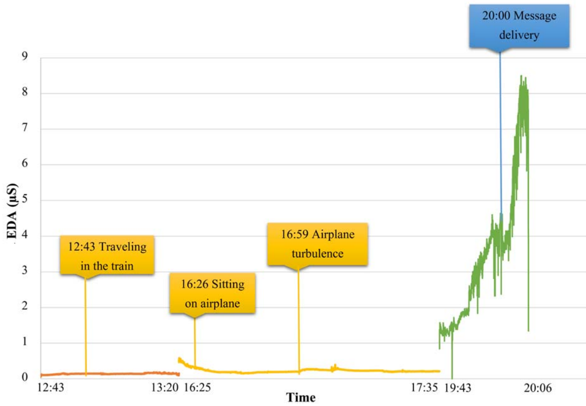

Figure 2: Activities of the subject during the day one.

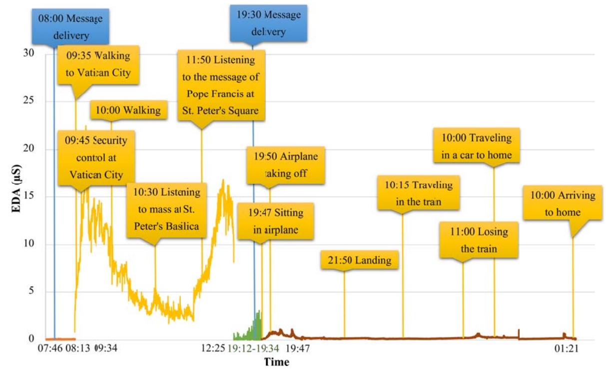

Figure 3: Activities of the subject during the day four.

shown in Table 1, the subject was stressed about 25% of the total time in the which the data was collected during day two (3.5h).

3.4.3 Day three. During the first message, the subject was walking from hotel to bus station, this message was delivered at 08:00 with high volume and of authority level, as the subject was walking on the street (outdoors), the EDA signal was influenced by the hot weather (around 25 °C). Around 11:00 hours, the subject took a walking tour for visiting different Roman architectures. Nearby 12:40 hours, user had to pass a security control at *Pantheon*. Then, during the afternoon subject walked around *Pantheon*. The second message was delivered at 20:00 hours, it was with a high volume and authority level; in that moment subject was resting at the hotel. During the third day, the subject was monitored about 10 hours. We can see the subject was stressed about 13% of the total time being monitored.

3.4.4 Day four. As show in Figure 3, the EDA values were the minimum despite the first message (configured with a medium volume and consensus level) was delivered while the subject was having breakfast at the hotel (his physical activity was quiet). The second message was delivered thirty minutes before the usual time because the subject had to board his return flight at Fiumicino Airport (Rome). During this message, the subject was sitting on the departure lounge of the airport. The stress detector did not recognize any stress episode, and it was also consistent with a not-stress self-report.

Figure 3 details different subject's activities, in the morning he visited the Vatican City, and he participated in Catholic activities. We found that about 16% of the total time (9.6 hours of data monitoring) correspond to stress episodes (see Table 1), which overlapped with the security control for entering the Vatican City and with the moment when he missed the train to go back home.

3.5 Identified emotional triggers

In real-world conditions, there exist many stimuli, either external or internal, that might trigger a range of emotions in users, which we call emotional triggers. In this case study, we relied on multiple sources in order to identify the emotional triggers that were observed, which can be classified as proposed by [25] into *environmental*, *physical* and *interaction* emotional triggers among others.

Figure 4 shows a pie graph resulting from the event analysis of the subject's activities classified according to the type of emotional triggers. The most frequent was the environmental type with a 42% (*i.e.,* the hot weather). However, the less frequent emotional trigger was that related to *"interaction with others"* (*e.g.,* interviews at security controls). Regarding the activities: traveling in bus or train, waiting in queues of landmarks were identified as other emotional trigger (24%).

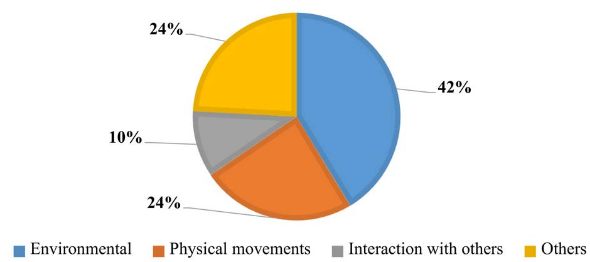

Figure 4: Classification of the tourist activities according to the emotional trigger classification of [25].

Regarding the most influential environmental emotional trigger, it was the high temperatures during the four days in Rome. This typically produces higher sweating on users [5], which also explains why the collected EDA values were greater than 10 μS. This external threat is challenging as it influences the EDA measurement; and this is why choosing a robust approach relying on variations rather than absolute values is important. No other special weather events were remarkable to report.

Some other potential threats or factors affecting user emotional state would fall within the category of perception emotional triggers. For instance, the lack of awareness on the utility of certain devices or on the usefulness of the mobile app might generate stress as well. Besides, the usage of technologies (*i.e.,* the smartphone and the E4-wristband) could be relevant if the continuous usage would cause a negative emotion (or any sort of user dissatisfaction). However, such sort of emotional triggers were not observed, being confirmed by the post hoc interviews, in which the subject reported that the E4-wristband did not cause any discomfort, and possibly also mitigated by the familiarity of the user with smartphone apps.

Additionally, tourism and vacations are moments to escape from the daily stress of work. However, some user activities or situations can generate stress in travelers too. Waterhouse et al. [49] argue that travelers reported "going to an airport" causes stress. For instance, the subject reported some episodes of stress when the airplane taking off and landing. Other potential issue is the inability to speak in the native language of the destination country, which has also been studied in different works [6, 49]. In our case, the end-user was not able to speak in Italian, which can be an issue during the security controls of museums because the subject could not communicate with the police. The identification of emotional triggers is key in order to assess the potential threats that may affect the subject emotional state. In line with this notion, we obtained a dataset with temporal event annotations, which can be used for understanding the user context and answering our research question as discussed in the following.

4 RESULTS

Table 2 summarizes when the messages were delivered along with other contextual information (*i.e.,* user activity, place, temperature of weather, and other sensor information such as human body temperature and accelerometers). We can see that the subject reported stress when the first three messages were delivered on day one and two, which is consistent with the outcome of the detector. This stress might have been caused by the settings of the software service (*i.e.,* high/medium volume and consensus level) since no other external emotional trigger was reported to happen at the same time. These three episodes of stress are therefore cataloged as actionable emotions.

However, a particular case was presented on day three (second message), where the user reported no-stress but the stress detector detector did. This mismatch could be produced due to the hot weather (28 °C); consequently, influencing on the EDA signals. The lack of context awareness by the stress detector would provide high levels of uncertainty to understand these cases. Thus, our hypothesis to use emotions from physiological data as implicit user feedback would be supported only if it can be accompanied by a reliable analysis and understanding of the user context.

4.1 Designing a Context Analyzer

The previous observations and results highlight the necessity for the stress detector to properly understand how the context influences on the user emotions. Thus, we argue context information can be handled by a *context analyzer* governed by a *rule base* that allows the system to determine whether the detected stress can be considered as an actionable emotion [9] and consequently as implicit user feedback. A context analyzer should model the user context, processing a massive amount of data from different sensors surrounding the user. To this end, it can firstly determine emotional triggers (*e.g.,* uncomfortable or comfortable temperature) and then classify them based on the taxonomy by Kanjo et al. [25]. In order to understand whether the recognition of negative emotions (physiological stress) can be used as implicit user feedback, we consider that a context analyzer should inspect the following three types of emotional triggers: environmental noise, social interaction with others, and physical movements. As collected in the case study, the presence of environmental noise can be sensed by the smartphone's microphone whereas the physical movement can rely on accelerometers. For sensing social interaction, several sophisticated alternatives can be considered depending on the situation and the sensors available. For instance, the camera would allow for checking the presence of multiple people in front of the field view while the microphone would facilitate checking whether they are engaged in an ordinary conversation if the decibel level is between 40dB and 80dB [12].

The rule base would work considering three primary cases for the assessment of the actionable emotions:

- **Case**$_1$, when the software service is delivered **AND** the emotion detector detects an emotion (TRUE) **AND** the context analyzer does not detect any emotional trigger **THEN** the emotion is actionable.
- **Case**$_2$, when the context analyzer detects one emotional triggers **AND** the emotion detector detects an emotion **AND** the software service was delivered **THEN** the result is an undetermined status.
- **Case**$_3$, when the context analyzer detects an emotional trigger **AND** the emotion detector detects an emotion **AND** the software service was not delivered **THEN** the result is a non-actionable emotion detected.

Notice that the case$_2$ results undetermined because, with the available information, it could be a confounded consequence of both the service and the emotional trigger. Table 3 summarizes the three cases for the assessment of the actionable emotions.

5 THREATS TO VALIDITY

The main threats to validity concern internal, construct and external validity [41].

5.1 Internal validity

It concerns on *additional factors* that may affect an observed variables. In our study, one of these threats is concerned about data quality, which could have been affected due to certain aspects like tracking loss, sensitivity of the wearable device to the ambient temperature. We reduce this threat, by filtering the noise from the collected physiological data [31]. Another threat might be the fact that the subject were affected by the presence of an observer along the four-days study. To mitigate this threat, we decided to invite a subject that was a close friend of the observer. As the subject is exposed to the treatment several times, the subject responses (stress state) might be affected as the time passes and the stimulus is repeated. However, the software service is delivered with a difference of 12 hours, therefore we consider this threat is not very critical.

5.2 Construct validity

It refers to *the degree a test measures what it claims to be measuring*. The use of a single device to measure physiological stress (a

Table 2: Summary of the Contextual information when the messages were delivered.

Day	Moment	Volume	Persuasiveness level	User activity	Place	User self-report	Stress detector	Other external context information
1	**Night** (19:43-20:06)	High	Authority	Subject was resting in his hotel room after to arrive to Rome.	Indoor	Stressed	Stress detected	27 °C-Passing clouds. Physical activity: quiet (rest). Skin temperature: 36.2 °C.
2	**Morning** (07:43-08:06)	High	Consensus	He was at hotel getting ready to go out (walking around the room).	Indoor	Stressed	Stress detected	20 °C Sunny. Physical activity: Walking. Skin temperature: 36.0 °C.
	Night (19:43-20:13)	Medium	Consensus	Subject was having dinner at Rome restaurant.	Outdoor	Stressed	Stress detected	27 °C Sunny. Physical activity: quiet (sitting). Skin temperature: 36.5 °C.
3	**Morning** (07:46-08:06)	High	Authority	He was walking from hotel to bus station.	Outdoor	Not stressed	Not stress detected	22 °C Passing clouds. Physical activity: walking. Skin temperature: 36.4 °C.
	Night (19:50-20:20)	High	Authority	He was resting at hotel after to visit the Pantheon.	Indoor	Not Stressed	Stress detected	28 °C Passing clouds. Physical activity: quiet (sitting). Skin temperature: 36.1 °C.
4	**Morning** (07:46-08:13)	Medium	Consensus	He was having breakfast at Hotel.	Indoor	Not stressed	Not stress detected	23 °C Sunny. Physical activity: quiet (sitting). Skin temperature: 36.2 °C.
	Night (19:12-19:34)	High	Authority	He was sitting at Rome Airport waiting for his flight.	Indoor	Not stressed	Not stress detected	29 °C Passing clouds. Physical activity: Quiet (sitting). Skin temperature: 36.4 °C.

Table 3: Possible combinations of internal context and external context variables for the assessment of the actionable emotion.

		Emotional trigger (stressor)				
	Software service	**Environmental noise**	**Interacting with others**	**Physical movements**	**Stress detector**	**Result**
Case$_1$	Delivered	Non-Detected	Non-Detected	Non-Detected	True	Actionable emotion
Case$_2$	Delivered	One or more triggers detected			True	Undetermined
Case$_3$	Non-Delivered	Any trigger detected			True	Emotion detected

construct) as well as a self-reported stress (yes/no questions) could be considered as threats to construct validity of this study. This threat was reduced by involving an observer to verify the presence of potential emotional triggers from the context.

5.3 External validity

It concerns the *generalization* of the findings beyond the validation settings. Since we employed the case study research method, the findings are bound by the scope of the selected case. In this paper we do not target to generalize our results but to remark the importance to analyze the context of the users in order to decide whether physiological data can be used as implicit user feedback in future designs. In this context, the setting of our study (wearable sensors, observer, context analyzer) was prepared for getting a better understanding on the conditions in which the different episodes of stress were experienced by our case (a tourist).

6 CONCLUSIONS

The main goal of the present case study has been to investigate whether physiological data can be considered as implicit user feedback. As the case study has been conducted in real-life conditions, different factors and tourist activities have been identified and classified as emotional triggers (*i.e.,* environmental, physical movements, interaction with others). For the identification of such emotional triggers we realize the importance of having a context analyzer capable of processing multisensorial as well as physiological data. Thus, our research goes beyond the mere detection of raw emotions from physiological data, by combining emotions and user context data in order to determine whether the emotion is actionable and it can therefore be intended as implicit user feedback. This means resourceful information that can be used in the continuous maintenance and improvement of software quality based on negative user experience [10].

Further work is needed to implement the designed context analyzer. We also plan to investigate in depth how implicit user feedback can be used effectively for reconfiguring software services in different context of use.

ACKNOWLEDGMENT

Special thanks go to the volunteer subject in the case study. This work has been supported by the projects: KUSISQA (014-2019-FONDECYT-BM-INC.INV) from the National Fund for Scientific and Technological Development (FONDECYT-PERU) and World Bank, and Datos 4.0 (TIN2016-78011-C4-1-R) supported by MINECO-AEI/FEDER-UE. Alejandro Catala is a *Juan de la Cierva* Researcher (IJC2018-037522-I). This research is partly funded by the Spanish Ministry of Science, Innovation and Universities (grant RTI2018-099646-B-I00) and the Galician Ministry of Education, University and Professional Training (grant ED431G2019/04). Some of the previous grants were co-funded by the European Regional Development Fund (ERDF/FEDER program).

REFERENCES

[1] Tamara Adlin and John Pruitt. 2010. *The Essential Persona Lifecycle: Your Guide to Building and Using Personas.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[2] Louis Atallah, Benny Lo, Rachel King, and Guang-Zhong Yang. 2010. Sensor Placement for Activity Detection Using Wearable Accelerometers. In *2010 International Conference on Body Sensor Networks*. IEEE.

[3] Franck Berthelon and Peter Sander. 2013. Emotion ontology for context awareness. In *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*. Institute of Electrical and Electronics Engineers (IEEE).

[4] Albert Bifet and Ricard Gavaldà. 2007. Learning from Time-Changing Data with Adaptive Windowing. In *Proceedings of the 2007 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 443–448.

[5] Wolfram Boucsein. 2012. *Electrodermal Activity.* Springer US.

[6] Chun-Chu Chen, James F. Petrick, and Moji Shahvali. 2014. Tourism Experiences as a Stress Reliever. *Journal of Travel Research* 55, 2 (aug 2014), 150–160.

[7] Robert Cialdini. 1993. *Influence: Science and Practice.*

[8] Mark Claypool, Phong Le, Makoto Wased, and David Brown. 2001. Implicit Interest Indicators. In *Proceedings of the 6th International Conference on Intelligent User Interfaces* (Santa Fe, New Mexico, USA) *(IUI '01)*. ACM, New York, NY, USA, 33–40.

[9] Nelly Condori-Fernandez. 2017. HAPPYNESS: An Emotion-aware QoS Assurance Framework for Enhancing User Experience. In *Proceedings of the 39th International Conference on Software Engineering Companion* (Buenos Aires, Argentina) *(ICSE-C '17)*. IEEE Press, Piscataway, NJ, USA, 235–237.

[10] Nelly Condori-Fernandez, Joao Araujo, Alejandro Catala, and Patricia Lago. 2020. Towards a Non-Functional Requirements Discovery Approach for Persuasive Systems. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing* (Brno, Czech Republic) *(SAC '20)*. Association for Computing Machinery, New York, NY, USA, 1418–1420.

[11] Nelly Condori-Fernandez and Franci Suni Lopez. 2017. Using Emotions to Empower the Self-adaptation Capability of Software Services. In *Proceedings of the 2Nd International Workshop on Emotion Awareness in Software Engineering* (Buenos Aires, Argentina) *(SEmotion '17)*. IEEE Press, Piscataway, NJ, USA, 15–21.

[12] Don Davis and Carolyn Davis. 1997. *Sound System Engineering.* Focal Press.

[13] Beat Fasel and Juergen Luettin. 2003. Automatic facial expression analysis: a survey. *Pattern Recognition* 36, 1 (jan 2003), 259–275.

[14] Enrique Garcia-Ceja, Venet Osmani, and Oscar Mayora. 2016. Automatic Stress Detection in Working Environments From Smartphones' Accelerometer Data: A First Step. *IEEE Journal of Biomedical and Health Informatics* 20, 4 (July 2016), 1053–1060.

[15] Daniela Girardi, Filippo Lanubile, and Nicole Novielli. 2017. Emotion detection using noninvasive low cost sensors. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE.

[16] Michael Godfrey and Daniel German. 2008. The past, present, and future of software evolution. In *2008 Frontiers of Software Maintenance*. IEEE.

[17] Zied Guendil, Zied Lachiri, Choubeila Maaoui, and Alain Pruski. 2015. Emotion recognition from physiological signals using fusion of wavelet based features. In *2015 7th International Conference on Modelling, Identification and Control (ICMIC)*. IEEE.

[18] Kun Han, Dong Yu, and Ivan Tashev. 2014. Speech emotion recognition using deep neural network and extreme learning machine. In *INTERSPEECH.*

[19] Jennifer Healey and Rosalind W. Picard. 2005. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE Transactions on Intelligent Transportation Systems* 6, 2 (June 2005), 156–166.

[20] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *2008 Eighth IEEE International Conference on Data Mining*. IEEE.

[21] Paola Andrea Jaramillo Garcia, Luis Ignacio Lopera Gonzalez, and Oliver Amft. 2015. Using implicit user feedback to balance energy consumption and user comfort of proximity-controlled computer screens. *Journal of Ambient Intelligence and Humanized Computing* 6, 2 (01 Apr 2015), 207–221.

[22] Gawesh Jawaheer, Martin Szomszor, and Patty Kostkova. 2010. Comparison of Implicit and Explicit Feedback from an Online Music Recommendation Service. In *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems* (Barcelona, Spain) *(HetRec '10)*. ACM, New York, NY, USA, 47–51.

[23] Gawesh Jawaheer, Peter Weller, and Patty Kostkova. 2014. Modeling User Preferences in Recommender Systems: A Classification Framework for Explicit and Implicit User Feedback. *ACM Trans. Interact. Intell. Syst.* 4, 2, Article 8 (June 2014), 26 pages.

[24] Oliver John and Sanjay Srivastava. 1999. The Big Five Trait taxonomy: History, measurement, and theoretical perspectives.

[25] Eiman Kanjo, Luluah Al-Husain, and Alan Chamberlain. 2015. Emotions in context: examining pervasive affective sensing systems, applications, and analyses. *Personal and Ubiquitous Computing* 19, 7 (01 Oct 2015), 1197–1212.

[26] Byoung Ko. 2018. A Brief Review of Facial Emotion Recognition Based on Visual Information. *Sensors* 18, 2 (jan 2018), 401.

[27] Tong Queue Lee and Young Park Yong-Tae Park. 2008. A time-based approach to effective recommender systems using implicit feedback. *Expert Systems with Applications* 34, 4 (may 2008), 3055–3062.

[28] Luis Leiva. 2011. Restyling Website Design via Touch-based Interactions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) *(MobileHCI '11)*. ACM, New York, NY, USA, 599–604.

[29] Gai Li and Qiang Chen. 2016. Exploiting Explicit and Implicit Feedback for Personalized Ranking. *Mathematical Problems in Engineering* 2016 (2016), 1–11.

[30] Nathan Liu, Evan Xiang, Min Zhao, and Qiang Yang. 2010. Unifying Explicit and Implicit Feedback for Collaborative Filtering. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (Toronto, ON, Canada) *(CIKM '10)*. ACM, New York, NY, USA, 1445–1448.

[31] Franci Suni Lopez, Nelly Condori-Fernandez, and Alejandro Catala. 2019. Towards Real-Time Automatic Stress Detection for Office Workplaces. In *Information Management and Big Data*. Springer International Publishing, 273–288.

[32] Walid Maalej, Maleknaz Nayebi, Timo Johann, and Guenther Ruhe. 2016. Toward Data-Driven Requirements Engineering. *IEEE Software* 33, 1 (jan 2016), 48–54.

[33] Nazim Madhavji, Juan Fernandez-Ramil, and Dewayne Perry. 2006. *Software Evolution and Feedback: Theory and Practice.* John Wiley & Sons, Inc., USA.

[34] Nesrine Mezhoudi and Jean Vanderdonckt. 2015. A user's feedback ontology for context-aware interaction. In *2015 2nd World Symposium on Web Applications and Networking (WSWAN)*. IEEE.

[35] Itzel Morales-Ramirez, Anna Perini, and Renata S.S. Guizzardi. 2015. An ontology of online user feedback in software engineering. *Applied Ontology* 10, 3-4 (Dec 2015), 297–330.

[36] Oscar Martinez Mozos, Virginia Sandulescu, Sally Andrews, David Ellis, Nicola Bellotto, Radu Dobrescu, and Jose Manuel Ferrandez. 2017. Stress Detection Using Wearable Physiological and Sociometric Sensors. *International Journal of Neural Systems* 27, 02 (mar 2017), 1650041.

[37] Cyrus H Nozad, L Madison Michael, D Betty Lew, and Christie F Michael. 2010. Non-allergic rhinitis: a case report and review. *Clinical and Molecular Allergy* 8, 1 (Feb. 2010).

[38] Tejaswini Panure and Shilpa Sonawani. 2019. Novel Approach for Stress Detection Using Smartphone and E4 Device. In *Sustainable Communication Networks and Application*. Springer International Publishing, 736–745.

[39] Ladislav Peska. 2016. Using the Context of User Feedback in Recommender Systems. In *Proceedings 11th Doctoral Workshop on Mathematical and Engineering Methods in Computer Science, MEMICS 2016, Telč, Czech Republic, 21st-23rd October 2016.* 1–12.

[40] Arkalgud Ramaprasad. 1983. On the definition of feedback. *Behavioral Science* 28 (1983), pp 4–13.

[41] Per Runeson and Martin Höst. 2009. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering* 14, 2 (2009), 131–164.

[42] Akane Sano and Rosalind W. Picard. 2013. Stress Recognition Using Wearable Sensors and Mobile Phones. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE.

[43] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In *Proceedings of the 2018 on International Conference on Multimodal Interaction - ICMI '18*. ACM Press.

[44] Pekka Siirtola. 2019. Continuous stress detection using the sensors of commercial smartwatch. In *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers - UbiComp/ISWC '19*. ACM Press.

[45] Yang Song and Li-wei He. 2010. Optimal Rare Query Suggestion with Implicit User Feedback. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA) *(WWW '10)*. ACM, New York, NY, USA, 901–910.

[46] Senthil Sriramprakash, Vadana Prasanna, and Ramana Murthy. 2017. Stress Detection in Working People. *Procedia Computer Science* 115 (2017), 359–366.

[47] Franci Suni Lopez and Nelly Condori-Fernandez. 2017. *Design of an Adaptive Persuasive Mobile Application for Stimulating the Medication Adherence*. Springer International Publishing, Cham, 99–105.

[48] Stefanos Vrochidis, Ioannis Kompatsiaris, and Ioannis Patras. 2011. Utilizing Implicit User Feedback to Improve Interactive Video Retrieval. *Advances in Multimedia* 2011 (2011), 1–18.

[49] James Waterhouse, Thomas Reilly, and Ben Edwards. 2004. The stress of travel. *Journal of Sports Sciences* 22, 10 (oct 2004), 946–966.

[50] Markus Zanker and Markus Jessenitschnig. 2009. Collaborative Feature-Combination Recommender Exploiting Explicit and Implicit User Feedback. In *2009 IEEE Conference on Commerce and Enterprise Computing*. IEEE.

[51] Jianhua Zhang, Zhong Yin, Peng Chen, and Stefano Nichele. 2020. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion* 59 (July 2020), 103–126.