Appraisal Theory-based Mobile App for Physiological Data Collection and Labelling in the Wild

Fanny Larradet
Istituto Italiano di Tecnologia, Genova, Italy
fanny.larradet@iit.it

Radoslaw Niewiadomski
Istituto Italiano di Tecnologia, Genova, Italy
radoslaw.niewiadomski@iit.it

Giacinto Barresi
Istituto Italiano di Tecnologia, Genova, Italy
giacinto.barresi@iit.it

Leonardo S. Mattos
Istituto Italiano di Tecnologia, Genova, Italy
leonardo.demattos@iit.it

ABSTRACT

Numerous studies on emotion recognition from physiological signals have been conducted in laboratory settings. However, differences in the data on emotions elicited in the lab and in the wild have been observed. Thus, there is a need for systems collecting and labelling emotion-related physiological data in ecological settings. This paper proposes a new solution to collect and label such data: an open-source mobile application (app) based on the appraisal theory. Our approach exploits a commercially available wearable physiological sensor connected to a smartphone. The app detects relevant events from the physiological data, and prompts the users to report their emotions using a questionnaire based on the Ortony, Clore and Collins (OCC) Model. We believe that the app can be used to collect emotional and physiological data in ecological settings and to ensure high quality of ground truth labels.

ACM Reference Format:

Fanny Larradet, Radoslaw Niewiadomski, Giacinto Barresi, and Leonardo S. Mattos. 2019. Appraisal Theory-based Mobile App for Physiological Data Collection and Labelling in the Wild. In *Adjunct Proceedings of the 2019 ACM International Joint conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC'19 Adjunct), September 9-13, 2019, London, United Kingdom*, ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3341162.3345595

UbiComp/ISWC '19 Adjunct, September 9-13, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6869-8/19/09…$15.00
https://doi.org/10.1145/3341162.3345595

1 INTRODUCTION

Several works have shown that physiological signals can constitute indices for automatic emotion recognition [1]. Differences were observed when comparing physiological data of emotions induced in the lab to real-life emotional reactions [2]. Difficulties in building the affect-related datasets in ecological settings, e.g., establishing the ground truth, are well documented in the literature [3, 4]. This paper proposes a novel system to collect and label physiological data in the wild. It is an open-source mobile application (app) based on the appraisal theory. The Ortony, Clore and Collins (OCC) model [5] has been chosen as it was successfully used in affective computing applications in the past [6, 7]. It predicts 22 emotion labels based on valence and the emotional trigger type (event, object or agent). The app detects additional heart rate to predict emotional events from physiological signals [8]. Once relevant events are detected, the app prompts the users to provide the appraisal evaluation of the event, helping them to define their emotional state. Unlike existing solutions, which often only use a constrained list of emotional labels [9] or dimensions [10], in our app we introduce a questionnaire based on appraisal theory to help the user provide the ground truth for his/her emotional states. By collecting the information about appraisal process we hope to improve the ground-truth labelling and to provide more consistent annotation of corresponding physiological signals.

This work is part of project TEEP-SLA, which aims at automatically detecting emotions from physiological signals for Amyotrophic Lateral Sclerosis (ALS) patients to improve their communicative abilities [11]. The long-term aim is to create a large dataset to be used in emotion recognition from physiological data collected in natural settings.

2 STATE OF THE ART

Emotion recognition from physiological signals

Emotion recognition from physiological data collected in the lab was often addressed [1]. Most of the studies use measurements of Heart Rate (HR), Skin Conductance (SC), Electro-Dermal Activity (EDA), Galvanic Skin Response (GSR), Skin

Temperature (ST), and Respiration. Fusions of several signals were also studied, e.g., the combination of HR, EDA and ST, also used in our work, has been studied in the past in [9] to classify anger, surprise, fear, frustration, and amusement with an average recognition rate of 83%.

Studies using data collected in ecological settings are rare, and most of them focus primarily on stress detection [12, 13, 14]. Some studies investigating affective states focused on moods [15] as they can be measured at any time of the day. It is more difficult to collect and label the data of emotions in ecological settings, as they are usually much shorter and more momentary than moods [16]. Therefore, methods which ask the user to report emotions at fixed time intervals, e.g., [13], might not be appropriate to collect such data. Other studies use mobile apps to collect self-reports in the wild (Section 2).

Methods for emotional self-reporting in the wild

According to [17], existing techniques for emotional state self-reporting can be divided into two groups: free response and fixed-response labelling. While the first group allows for a higher precision of labelling (custom labels [18], verbal reports [19]), it makes it difficult to develop machine learning recognition models due to a potentially wide range of emotion labels selected by users. Constrained solutions include the usage of a finite list of labels (e.g., [9]) or dimensional models such as valence-arousal (e.g., [4]) or pleasure-arousal-dominance (e.g., [20]). More user-friendly techniques may be used for reporting such as emoticons [21]. Affect dimensions are usually reported through the Self-Assessment Manikin (SAM) method [18] or through 2D point maps [10].

In [3], guidelines are provided for emotional labelling in the wild by comparing the results of different methods. A combination of manual reports and automatically triggered prompts is advised, as well as providing the means to the user to manually correct the timespan of an emotional event. Unlike [3], that used time-based trigger, we used prompting based on physiological cues [8] and we opted for an experimenter-free data gathering protocol. We reduced the role of the experimenter hoping that it may help different research teams to contribute in future to the creation of a large shared dataset.

Methods for emotional physiological data collection

To collect physiological signals of emotions in the laboratory, researchers use a great range of emotion induction techniques. However, they usually induce emotions of low intensity [2]. Some studies have been inducing stress using more ecologically valid procedures, e.g., putting subjects into stressful situations such as driving [22] or sky diving [23]. In real-life settings, the physiological data labelling and segmentation (i.e., defining the start and end of an emotion) are a much bigger challenge [4]. A few studies used mobile apps to collect both physiological data and affect related states. The most common ones collect stress levels [14, 19] or moods [10, 15] .

Healey and colleagues [4] conducted a real-life experiment using a mobile phone app to study different labelling methodologies for physiological data collection. They collected data and self-reports in the form of discrete labels and dimensional models (valence and arousal) and drew attention to some difficulties linked to self-reporting. For instance, from the reports, the label "anxious" was annotated both as a positive and negative emotion. This example highlights a need for a scheme to help users pick labels. They reached a rate of 85% for classifying arousal and 70% for classifying valence using GSR and HR on manually extracted data segments of various durations.

3 PRELIMINARY STUDY

In a preliminary study (PS) we collected physiological data in ecological settings using a standard paper-based self-reporting method. 4 subjects (3 males, 1 female; avg. age 29 years) participated in the study. The experimental procedures follow the IIT ADVR TEEP02 protocol, approved by the Ethical Committee of Liguria Region on September 19, 2017.

Study protocol

The subjects wore the Empatica E4 bracelet [24] for 5 days, 12 hours a day. They were asked to remove the bracelet at night, during sport and showers. They kept a hand-written journal of their emotions. We focused on the 3 most common basic states: happy, sad and angry. For each emotional event, participants were asked to report its start and end time as well as the intensity using a 5 point Likert scale.

Issues and lessons learned

We collected Blood Volume Pressure (BVP), EDA and ST data for a total of 234h 02m 29s. This pilot study gave us a great number of insights into the problems faced when collecting physiological data in ecological settings. It also confirmed the issues previously discussed in the literature e.g., [3, 4]. Several subjects forgot to wear the device and failed to report some relevant events. When the data was analyzed after the study, we asked some participants about moments in the day where the physiological signals seemed to be particularly different from the baseline. Only then they remembered the events which they had failed to report before. Additionally, some subjects forgot to rate the intensity of certain emotions.

Furthermore, it seems that our participants had difficulty with distinguishing what constitutes an emotion. For instance, an event "Happy: 8AM to 8PM intensity rating 1" was reported by a participant. However, the long duration and low intensity makes us believe that in this case the user was referring to a mood rather than an emotion [16].

4 THE PROPOSED SOLUTION

Collecting and labelling the physiological data of emotions in ecological settings brings many difficulties. In order to address them, we aimed to create a mobile application which:

(1) can be used to capture physiological signals of spontaneous emotions during every-day activities;
(2) is minimally intrusive;
(3) guides the user through the process of reporting relevant events, by acquiring the necessary information to infer the related affective states, and without asking the user to pick any emotional labels;
(4) helps the user to provide meaningful annotation by differentiating emotions from moods;
(5) detects the relevant events from the physiological data and prompts the user about it;
(6) provides a limited set of ground-truth labels to be used in recognition and classification models.

Taking into account the results of the preliminary study (Section 3) we proposed a solution based on appraisal theory, a commercially available physiological sensor, a mobile application, as well as a state-of-the-art event detection algorithm.

Self-reporting about relevant events

To fulfil the requirements 3, 4 and 6, we use appraisal theory for self-reporting which acquires the whole appraisal process around the event. The resulting annotation consists of a limited set of labels (single appraisals or emotional labels corresponding to a combinations of appraisals), and it can, therefore, be used to build classifiers with machine learning.

Unlike user-picked (UP) label-based datasets that use a specific set of labels for a specific application, exploiting appraisal theory to annotate the data allows one to build application-independent datasets. Indeed, the same dataset can be used in different application-specific recognition models, by choosing the relevant subset of emotional labels, or by detecting single appraisals. It provides for a greater information about the event (additional details on what led to the emotion) and a large number of labels to the experimenter without being cumbersome to the user since they do not need to choose such complex labels from a long list. Additionally, using appraisal theories allows for the creation of a single appraisal recognition model from physiological data [25]. Such models have rarely been studied so far but the results are promising [26]. The OCC model was chosen for its simplicity to create an adapted questionnaire comprehensible by non-experts.

Sensors

The Empatica E4 bracelet allowed us to fulfil requirements 1 and 2. This medical device was chosen for its sensors relevant to emotion detection: BVP, EDA and ST as well as kinematic data through a 3D accelerometer. Its small size allows for long duration experiments without being bothersome. The device comes with an API for mobile applications and an already processed BVP to Inter Beat Interval (IBI). Both raw BVP and calculated IBI are collected by the app to allow experimenters to perform their own peak detection method. The sensor has also been used in the past for research purposes [12].

The iPhone-based (iOS) mobile app use a Bluetooth connection to collect physiological data from the E4 bracelet.

The application modules

The emotion definition module. is designed to collect information about relevant emotional events. Using this module, the users first provide the duration of a relevant event. The maximum duration of the event was set to 5 minutes to limit the collection of moods as emotions are usually shorter. Next, they answer a series of questions according to the questionnaire (Fig. 1) and give the strength of the emotion.

To collect the information about the relevant events, we converted the OCC model into a question tree (see Fig. 1). For instance someone frightened by an incoming meeting would probably answer the example path in Fig. 1. We introduced small changes to the original model to differentiate mood from emotions. Indeed, according to [27], moods are *unconstrained in meaning*, while emotions are directed at specific objects, events or people. We therefore added a branch to the tree to provide the possibility to report such "unconstrained in meaning" experiences (see "Mood" branch in Fig. 1).

The event detection module. is used to detect relevant events from the data in real-time. The additional heart rate method [8] was used to detect relevant events and prompt the user to report his/her emotion at this time. It consists in detecting heart rate increases that are unrelated to activity (estimated using the accelerometer). Detected events create a *mandatory events* list, which is always accessible to the user on a separate tab of the app. By implementing this algorithm we fulfilled requirement 5 from the list presented in Section 4.

As the exact length of the detected event is unknown, we set it to the maximum time allowed for voluntary report: 5 minutes, 150s before and after the detected peak. The minimum time interval between two detected mandatory events is fixed at 1 hour to avoid life disturbance with too many prompts. If two or more events are detected within an hour, only the first event is added to the mandatory list and the remaining ones are ignored.

The Notification module. reminds the user to wear the device when needed and to report the events from the mandatory event list, if any. Reminders, when needed, are done at a rate of once every 15 min since emotional reports become less accurate as time passes [28]. When connection with the wristband is lost, notifications are prompted by the phone every 15 seconds until reconnection.

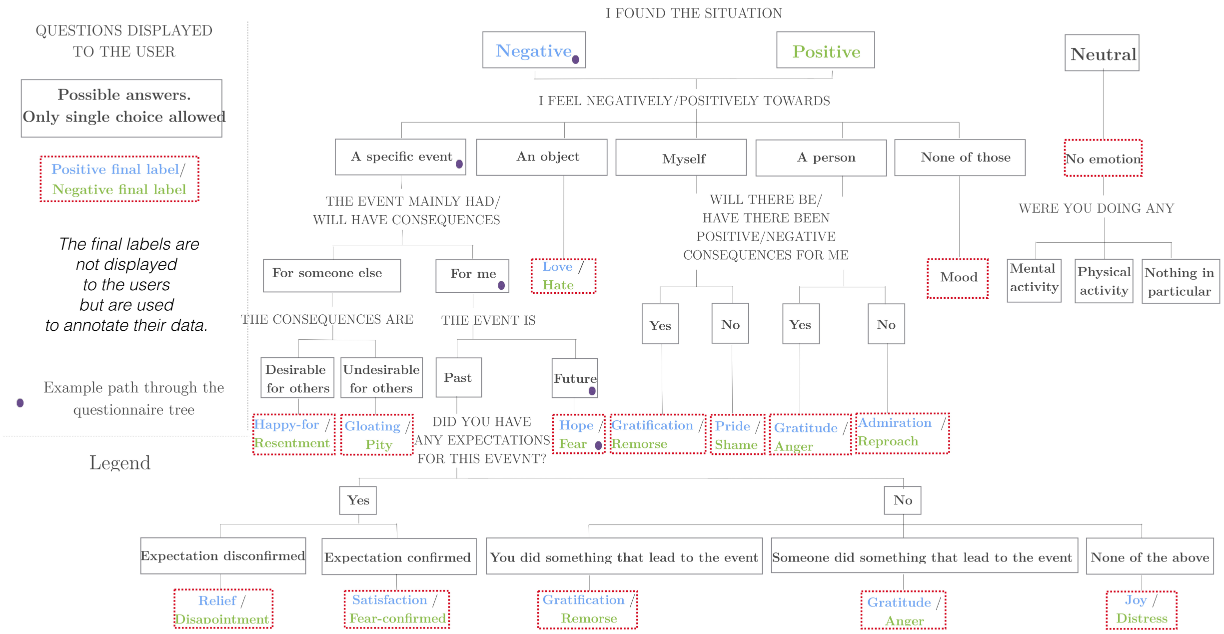

Figure 1: OCC-based questionnaire.

The application functionalities

The mobile application is separated into 5 tabs:

Voluntary reports. The users can voluntarily report an undetected event. They select the start and end time (max. duration of 5 min) then continue with the emotion definition module.

Mandatory event list tab. When an event is detected by the event detection module (section 4), a mandatory event is added to the list. An event will also be added if the E4 bracelet's button is pressed. Our PS highlighted that reporting the events as they happen may be difficult. However, referencing them later may decrease the time range precision. By pressing the button, the users manually add a new entry to the mandatory list with a precise timing (150s before and after the button press). They can then report the event later.

The remaining 3 tabs allow for a better experience with the app: to temporarily stop the notifications, check the battery level and visualize the reports using appropriate graphics. A video of the app is available in the supplementary materials.

5 DATA COLLECTION

We performed a data collection with 4 subjects (3 males, 1 female, avg. age 28 years) who wore the Empatica E4 bracelet and an iPhone 5C running the app for 5 days each. An additional question was added at the end of the appraisal tree

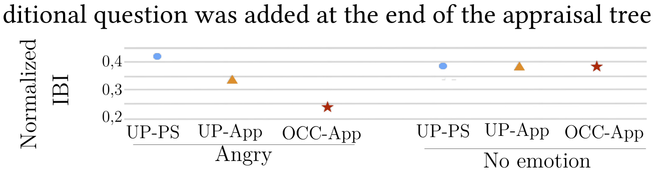

Figure 2: IBI averages for anger and baseline for each dataset.

where the users were asked to choose an emotional label between "happy", "sad", "angry" and "no emotion" (UP labels). 65% of the automatic prompts were rated as emotions, which suggest the suitability of the event detection. Additionally, some OCC labels were associated to both "angry" and "sad" UP labels. This highlights the shortcomings of the UP choice list to report emotional states. Finally, while HR is known to rise during anger events (lower IBI) compared to baseline [29], the normalized IBI average (aIBI) from anger events in the PS (0.41) is higher that the aIBI for no emotion (NE) (0.39) (Fig. 2) which is not consistent with literature and might indicate a poor quality dataset. While the aIBI from the UP anger events collected with the app (0.34) is lower than the one from the NE period (0.39), they are still very similar. The aIBI during OCC-labelled anger collected with the app are much lower (0.24) than the NE (0.39), which is consistent with literature and supports our hypothesis that this app allows for the collection of valuable emotional labels.

6 CONCLUSION AND FUTURE WORK

In this paper we proposed a new tool for collecting and labelling physiological signals, acquired during relevant events in ecological settings. This solution was designed according to appraisal theories, allowing the user to self-report the whole appraisal process around relevant events. The system is able to prompt the user to report the emotional self-assessment. We believe that the possibility of precise selection of relevant events timing and duration, the assistance given to the user to differentiate moods from emotions and the ability to report appraisals instead of emotion labels will improve the quality of the dataset compared to standard paper-type collection. To our knowledge, this is the first app for emotion reporting based on appraisal theory. It is open-source [30] and can be used by other researchers to extend the existing dataset. It provides the novel methodology to evaluate the physiological data collection of emotions in the wild. It allows to collect application-independent dataset containing an increased number of information about the emotional trigger and a great variability of label without it being cumbersome for the user. The same dataset might be used in the future, e.g., to create different application-specific classifiers, by choosing the relevant subset of the appraisals and emotional labels.

Future works include using this mobile application to collect data from a great number of subjects to create an open-source database. To better understand the potential benefits of using the OCC model, we will perform a study in which the users will report relevant events using the OCC-derived questionnaire as well as a list containing all 22 emotion labels.

ACKNOWLEDGMENT

The authors wish to thanks Sebastien Loustau and lumenAI. This work was partially funded by Fondazione Roma.

REFERENCES

[1] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, X. Yang, A review of emotion recognition using physiological signals, Sensors 18 (7) (2018) 1–41.
[2] F. H. Wilhelm, P. Grossman, Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment, Biological psychology 84 (3) (2010) 552–569.
[3] P. Schmidt, A. Reiss, R. Dürichen, K. Van Laerhoven, Labelling affective states in the wild: Practical guidelines and lessons learned, in: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, ACM, 2018, pp. 654–659.
[4] J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, M. Morris, Out of the lab and into the fray: towards modeling emotion in everyday life, in: International Conference on Pervasive Computing, Springer, 2010, pp. 156–173.
[5] A. Ortony, G. L. Clore, A. Collins, The cognitive structure of emotions, Cambridge university press, 1990.
[6] C. Conati, Probabilistic assessment of user's emotions in educational games, Applied artificial intelligence 16 (7-8) (2002) 555–575.
[7] C. Bartneck, Integrating the occ model of emotions in embodied characters, in: Workshop on Virtual Conversational Characters, Citeseer, 2002, pp. 39–48.
[8] M. Myrtek, G. Brügner, Perception of emotions in everyday life: studies with patients and normals, Biological psychology 42 (1-2) (1996) 147–164.
[9] F. Nasoz, K. Alvarez, C. L. Lisetti, N. Finkelstein, Emotion recognition from physiological signals using wireless sensors for presence technologies, Cognition, Technology & Work 6 (1) (2004) 4–14.
[10] E. A. Carroll, M. Czerwinski, A. Roseway, A. Kapoor, P. Johns, K. Rowan, M. Schraefel, Food and mood: Just-in-time support for emotional eating, in: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, IEEE, 2013, pp. 252–257.
[11] D. Beukelman, S. Fager, A. Nordness, Communication support for people with als, Neurology Research International 2011 (2011) 1–6.
[12] M. Gjoreski, M. Luštrek, M. Gams, H. Gjoreski, Monitoring stress with a wrist device using context, Journal of biomedical informatics 73 (2017) 159–170.
[13] K. Plarre, A. Raij, S. M. Hossain, A. A. Ali, M. Nakajima, M. Al'Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, et al., Continuous inference of psychological stress from sensory measurements collected in the natural environment, in: Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks, IEEE, 2011, pp. 97–108.
[14] K. Hovsepian, M. alAbsi, E. Ertin, T. Kamarck, M. Nakajima, S. Kumar, cstress: towards a gold standard for continuous stress assessment in the mobile environment, in: Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing, ACM, 2015, pp. 493–504.
[15] A. Zenonos, A. Khan, G. Kalogridis, S. Vatsikas, T. Lewis, M. Sooriyabandara, Healthyoffice: Mood recognition at work using smartphones and wearable sensors, in: 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), IEEE, 2016, pp. 1–6.
[16] E. K. Gray, D. Watson, R. Payne, C. Cooper, Emotion, mood, and temperament: Similarities, differences, and a synthesis, Emotions at work: Theory, research and applications for management (2001) 21–43.
[17] K. R. Scherer, What are emotions? and how can they be measured?, Social science information 44 (4) (2005) 695–729.
[18] M. Isomursu, M. Tähti, S. Väinämö, K. Kuutti, Experimental evaluation of five methods for collecting emotions in field settings with mobile applications, International Journal of Human-Computer Studies 65 (4) (2007) 404–418.
[19] A. Muaremi, B. Arnrich, G. Tröster, Towards measuring stress with smartphones and wearable devices during workday and sleep, BioNanoScience 3 (2) (2013) 172–183.
[20] R. Kocielnik, N. Sidorova, F. M. Maggi, M. Ouwerkerk, J. H. Westerink, Smart technologies for long-term stress monitoring at work, in: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, IEEE, 2013, pp. 53–58.
[21] A. Meschtscherjakov, A. Weiss, T. Scherndl, Utilizing emoticons on mobile devices within esm studies to measure emotions in the field, Proc. MME in conjunction with MobileHCI 9 (2009) 3361–3366.
[22] C. Dobbins, S. Fairclough, A mobile lifelogging platform to measure anxiety and anger during real-life driving, in: 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE, 2017, pp. 327–332.
[23] G. N. Dikecligil, L. R. Mujica-Parodi, Ambulatory and challenge-associated heart rate variability measures predict cardiac responses to real-world acute emotional stress, Biological psychiatry 67 (12) (2010) 1185–1190.
[24] Empatica, empatica, www.empatica.com (accessed 9 july 2019).
[25] M. Mortillaro, B. Meuleman, K. R. Scherer, Advocating a componential appraisal model to guide emotion recognition, International Journal of Synthetic Emotions (IJSE) 3 (1) (2012) 18–32.
[26] C. Smith, Dimensions of appraisal and physiological response in emotion, Journal of Personality and Social Psychology (1989) 339–353.
[27] G. L. Clore, A. Ortony, Psychological construction in the occ model of emotion, Emotion Review 5 (4) (2013) 335–343.
[28] I. B. Mauss, M. D. Robinson, Measures of emotion: A review, Cognition and emotion 23 (2) (2009) 209–237.
[29] G. E. Schwartz, D. A. Weinberger, J. A. Singer, Cardiovascular differentiation of happiness, sadness, anger, and fear following imagery and exercise., Psychosomatic medicine 43 (4) (1981) 343–364.
[30] F. Larradet, Mafed, https://gitlab.com/flarradet/mafed (accessed 9 july 2019).