Guidelines for Mobile Emotion Measurement

Joris H. Janssen
Dept. of Human Technology Interaction,
Eindhoven University of Technology &
User Experience Group, Philips Research
P.O. Box 513, 6500 MB
Eindhoven, The Netherlands
j.h.janssen@tue.nl

Egon L. van den Broek
Center for Telematics and Information
Technology (CTIT), University of Twente
P.O. Box 217, 7500 AE
Enschede, The Netherlands
vandenbroek@acm.org

ABSTRACT

Mobile emotion measurement (MEM) through physiological signals is a promising tool for both experiments and application. We provide 1) an overview of unobtrusive physiological sensors and 2) a review of studies that have tried to infer emotions from physiological signals. This review shows that there is a lack of general standards, low accuracy, and a doubtful validity of the results. To overcome these problems, we provide three guidelines for future research on MEM: validation, triangulation, and a physiology-driven approach. These guidelines enable the embedding of MEM in various professional and consumer settings, as a key factor in our every day life.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*Human factors*; J.4 [**Social and Behavioral Sciences**]: [Psychology]; J.7 [**Computers in Other Systems**]: [Consumer products]; I.2.m [**Miscellaneous**]: []

General Terms

Experimentation, Human Factors, Measurement, Performance, Reliability, Standardization

Keywords

Emotion, Physiology, Wearable, Affective computing, Physiological computing

1. INTRODUCTION

Would it not be great if a computer could warn us when we are under too much stress, if a tutoring system could monitor a student's attention, or if music could automatically be selected based on how we feel? These are typical examples of the next step in machine intelligence, which require an unobtrusive method for measuring one's mental state [19].

One of the promising ways of measuring these mental states is through physiological signals. The physiological counter-

parts of psychological phenomena have been researched for over a century. This has resulted in an enormous body of literature that describes physiological responses to all kinds of psychological states: mental workload, attention, pain, emotions, and dreams, to name but a few [4]. However, findings of studies manipulating psychological states and measuring physiological signals are typically inconsistent [10]. Nonetheless, the amount of research employing physiological signals with machine learning tools to predict mental states has exploded in the last decade; see also Table 2. Together with technological advances in physiological sensors, this leads the way to true mobile emotion measurement (MEM).

MEM would be of great benefit for mobile HCI. In the first place, it would enable real-time unobtrusive objective emotion measurements in experimental settings. This has advantages over subjective methods like questionnaires, which can only be done post-hoc and are very obtrusive. Additionally, it is doubtful whether subjective emotion reports always reflect the actual emotion. In contrast to physiological measurements, they are not free from social masking.

Currently, several wearable devices are being developed that can conduct physiological measurements in a unobtrusive, real-time fashion. This enables physiology-based MEM, as opposed to facial affect recognition that cannot be done through wearable devices. Moreover, they can be used anytime, where emotional speech processing only works in situations where one is speaking [19]. Hence, for studies using a mobile setting, emotions are best captured by physiological signals. Furthermore, MEM has not only methodological advantages, it also provides numerous opportunities for mobile applications. Possible applications include an affective mp3 player, continuous emotion communication, atmosphere creation, or even emotional jewelry. In turn, this might prove essential in realizing true ambient intelligence and forms next step in wearable and mobile computing [19].

In this paper, we will first provide an overview of mobile devices that measure physiology. In addition, we give an overview of state-of-the-art emotion prediction from physiological signals. We show that this prediction has not reached a satisfying level for MEM. Furthermore, we identify several methodological shortcomings of the studies done so far. To overcome these problems, we provide three guidelines that will help to further the development of successful MEM.

Table 1: General concerns with mobile emotion measurement (MEM).

1) Affective signals are typically derived through non-invasive methods to determine changes in physiology and, as such, are indirect measures. Hence, a delay between the actual change in emotional state and the recorded change in signal has to be taken into account, especially with mobile measurements.

2) Mobile measurements make physiological sensors sensitive to movement artifacts and differences in bodily position.

3) Most sensors are obtrusive, preventing their integration in real world applications.

4) Affective signals are influenced by (the interaction among) a variety of factors [4]. Some of these sources are located internally (e.g., a thought) and some are among the broad range of possible external factors (e.g., a signal outside). This makes affective signals inherently noisy, which is prominent in mobile measurements in real world environments.

5) Physiological changes can evolve in a matter of milliseconds, seconds, minutes or even longer. Some changes hold for only a brief moment, while others can even be permanent. Although seldom reported, the expected time windows of change are of interest [19]. In particular since changes can add to each other, even when having a different origin, as is often the case with mobile measurements.

6) Affective signals have large individual differences. This calls for methods and models tailored to the individual. It has been shown that personal approaches increase the performance of affect recognition; e.g., [3].

2. THE STATE-OF-THE-ART

A broad range of affective signals are used in affective sciences. Over the last decade, several unobtrusive devices for the processing of such signals have been developed. For instance, sensors for heart rate measurements have been integrated into a chair [1]. Furthermore, Healey and Picard [6] integrated heart rate and skin conductance measurements into a car, Ark et al. [2] have developed skin conductance sensors in a mouse, and Paulos et al. [13] combined wearable wireless heart rate sensors in a wrist band.

For a broad range of techniques, applications, and discussions concerning unobtrusive emotion measurement, we also refer to [20]. This shows the wide variety of possibilities for integrating physiological sensors in all kinds of everyday objects in an unobtrusive manner. After the signals have been captured, they have to be processed in real-time. When processing such signals, some general issues have to be taken in consideration, as are denoted in Table 1.

Typically, studies attempting to predict mental states using physiological signals conduct an experiment in which participants are brought into distinct mental states. A wide range of physiological signals are monitored from which numerous features are extracted. After feature extraction, machine learning techniques are employed to see if correct mental state predictions can be made based on the extracted features. See Table 2 for a concise review of such studies.

As illustrated by Table 2, a variety of physiological signals and machine learning techniques have been explored. Nonetheless, both the recognition performance and the number of emotions that the classifiers were able to discriminate are disappointing. Moreover, comparing the different studies is problematic because of the different settings the research was applied in, ranging from controlled lab studies to real world testing, the type of emotion triggers used, the number of target states to be discriminated, and the signals and features employed. Moreover, the general concerns are often disregarded, as denoted in Table 1. To conclude, there is a lack of standards, low prediction accuracy, and inconsistent results. For MEM to come to fruition, it is eminent to start dealing with these issues. This illustrates the need for a set of guidelines for MEM, as is done in the next section.

3. GUIDELINES

We identify three guidelines for MEM: 1) validity, 2) triangulation of measurements, and 3) a physiology-driven approach.

3.1 Validity

In the pursuit to trigger emotions in a more or less controlled manner, a range of methods have been applied: actors, images (IAPS), sounds (e.g., music), (fragments of) movies, speech, commercials, games, agents / serious gaming / virtual reality, real world experiences, and reliving of emotions. However, how to know which of these methods actually triggered participants' true emotions? This is a typical concern of validity, which is a crucial issue for MEM. Validity can be best obtained through four approaches: content, criteria-related, construct, and ecological validation, as we will discuss in this section.

Content validity refers to a) The agreement of experts on the domain of interest; e.g., limited to a specific application or group of patients; b) The degree to which a feature (or its parameters) of a given signal represents a construct; and c) The degree to which a set of features (or their parameters) of a given set of signals adequately represents all facets of the domain. For instance, employing only skin conductance level (SCL) will lead to a weak content validity when trying to measure emotion, as SCL is known to relate to the arousal component of an emotion, but not to the valence component. However, when trying to measure only emotional arousal, measuring only SCL may form strong content validity.

Criteria-related validity handles the quality of the translation from the preferred measurement to an alternative, rather than to what extend the measurement represents a construct. Emotions are preferably measured at the moment they occur, as is feasible with MEM. However, measurements before (predictive) or after (postdictive) the particular event are sometimes more feasible; e.g., through subjective questionnaires. The quality of these translations are referred to as predictive or postdictive validity. With emotion measurement, this is especially relevant for obtaining a reliable ground truth. The closer the ground truth measure to the actual emotion, the more reliable it becomes. A third form of criteria-related validity is concurrent validity: a metric for the reliability of measurements applied in relation to

Table 2: A summary of 11 studies that have tried to infer a mental state from physiological signals. They all employed a similar approach: first, a certain mental state (e.g., stress, certain emotions, mental workload) is induced in participants, while a number of physiological signals are measured. Subsequently, a variety of features is extracted and pattern recognition and machine learning techniques are employed to enable the automatic classification of the emotional states.

Source	Signals	Features	Selection/Reduction	Classifiers	Target	Result
[14]	$\mathcal{C},\mathcal{E},\mathcal{R},\mathcal{M}$	40	SFS, Fisher	LDA	8 emotions	81 %
[12]	$\mathcal{C},\mathcal{E},\mathcal{S}$	3		kNN, LDA	6 emotions	69 %
[16]	$\mathcal{C},\mathcal{E},\mathcal{B}$	18		SVM	6 emotions	42 %
[9]	$\mathcal{C},\mathcal{E},\mathcal{S}$	10		SVM	3 emotions	78 %
[11]	$\mathcal{C},\mathcal{E},\mathcal{S}$	12		kNN, LDA, ANN	6 emotions	84 %
[5]	$\mathcal{G}$	3	PCA	ANN	4 emotions	90 %
[6]	$\mathcal{C},\mathcal{G},\mathcal{R},\mathcal{M}$	22	Fisher	LDA	3 stress levels	97 %
[15]	$\mathcal{C},\mathcal{G},\mathcal{S},\mathcal{M},\mathcal{P}$	46		kNN, SVM, RT, BN	3 emotions	85 %
[22]	$\mathcal{C},\mathcal{G},\mathcal{S},\mathcal{P}$	11		SVM	2 stress levels	90 %
[21]	$\mathcal{C},\mathcal{E}$	20	ANOVA	SVM, ANN	2 fun levels	70 %
[8]	$\mathcal{C},\mathcal{E},\mathcal{M},\mathcal{R}$	15		SVM, ANFIS	4 affect states	79 %
[18]	$\mathcal{E},\mathcal{M}$	10	ANOVA, PCA	kNN, SVM, ANN	4 emotions	61 %

Notes. $\mathcal{C}$: Cardiovascular activity; $\mathcal{E}$: Electrodermal Activity; $\mathcal{R}$: Respiration; $\mathcal{M}$: Electromyogram; $\mathcal{B}$: Electroencephalogram; $\mathcal{S}$: Skin temperature; $\mathcal{P}$: Pupil Diameter; ANN: Artificial Neural Network; RT: Regression Tree; BN: Bayesian Network; SVM: Support Vector Machine; LDA: Linear Discriminant Analysis; kNN: k Nearest Neighbors; ANFIS: Adaptive neuro-fuzzy inference system; PCA: Principal Component Analysis; SFS: Sequential Forward Selection.

the preferred standard. For instance, the more emotions are discriminated, the higher the concurrent validity.

A construct validation process aims to develop a nomological network, or possibly an ontology or semantic network, build around the construct of interest. Such a network requires theoretically grounded, observable, operational definitions of all constructs and the relations between them. Such a network aims to provide a verifiable theoretical framework. The lack of such a network is one of the most pregnant problems physiological emotion measurement is coping with. A frequently occurring mistake is that emotions are denoted, where moods (i.e., longer object-unrelated affective states with very different physiology) are meant. This is very relevant, as it is known that moods are accompanied by very different physiological patterns than emotions are. Moreover, different signals relate to different emotional properties. For instance, arousal is strongly related to skin conductance and valence is thought to be reflected by heart rate variability.

Ecological validity refers to the influence of the context on measurements. As emotions are easily contaminated by contextual factors, using a similar context as the intended application for initial learning is of vital importance. Hence, emotion measurements done in controlled laboratory settings, are poorly generalizable to real-world applications. Hence, the need for MEM to make longitudinal real-world studies possible is pressing.

3.2 Triangulation
We propose to adopt the principle of triangulation on MEM, as applied in social sciences and human-computer interaction. This may deal with the noisy physiological signals inherent to MEM. For example, movement artifacts and corruption due to other signals can be major problems.

Heath [7] defines triangulation as "the strategy of using mul-

tiple operationalizations of constructs to help separate the construct under consideration from other irrelevancies in the operationalization". Using this strategy provides several advantages: 1) Distinct signals can be used to validate each other; 2) Extrapolations can be made based on multiple data sets, providing more certainty. In turn, corrections can be made to errors in a result set that clearly defy from other results; and 3) More solid ground is obtained for the interpretation of signals, as multiple perspectives are used.

Triangulation was, for example, successfully employed by [3], who showed that combining physiological signals and facial expressions leads to better predictions than using one of them. Also, [19] showed that the combination of a physiological parameter (i.e., heart rate variability) and speech parameters can provide more robust emotion recognition than either of them separately. Hence, we advise to record multiple affective signals, as is facilitated through MEM. Moreover, qualitative and subjective measures should accompany the signals; e.g., questionnaires, video recordings, interviews, and Likert scales. Systematic, well controlled research exploring the plethora of possible affective signals should increase the grip on the meaning of these signals.

3.3 A physiology-driven approach
A third guideline stems from the idea that physiological emotion measurement can never be entirely based on psychological changes. As discussed, there are many factors outside one's affective state that contaminate affective signals. Beside validation and triangulation, a physiology-driven perspective could be taken to deal with this [17].

Instead of expressing goals of MEM directly in terms of affective states, they can often be stated in terms of the affective signals themselves. For instance, instead of inferring an air-traffic controller's stress level, thresholding skin conductance level might be sufficient. Note that there always remains an

interpretation in affective states. Then, the use of syntactic or structural pattern recognition should be explored. Its hierarchical approach to simplifying complex patterns in affective signals could be valuable for MEM.

4. CONCLUSION

This paper described MEM through physiological signals and explained the lack of its success. Next, three guidelines were introduced from which MEM is expected to benefit significantly: validation, the principle of triangulation, and a physiological-driven approach.

With the guidelines provided and the future's progress ahead, we envision embedding of MEM in various professional and consumer settings, as a key factor of our every day life; cf. [19]. MEM fits the ambient intelligence vision perfectly. Although combining wearable and intelligent devices into smart environments is a great challenge, we strongly believe this holds a great promise for future technology and lifestyle. Would it not be an appealing idea to live in an empathic surrounding that adapts to your mood and emotions, which can even calm you or can help you concentrate when required?

5. REFERENCES

[1] J. Anttonen and V. Surakka. Emotions and heart rate while sitting on a chair. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 491–499, Portland, Oregon, USA, April 02–07 2005. New York, NY, USA: ACM.

[2] W. Ark, D. C. Dryer, and D. J. Lu. The emotion mouse. In H. Bullinger and J. Ziegler, editors, *Human Computer Interaction: Ergonomics and User Interfaces*, pages 818–823, London, 1999. Lawrence Erlbaum.

[3] J. N. Bailenson, E. D. Pontikakis, I. B. Mauss, J. J. Gross, M. E. Jabon, C. A. Hutcherson, C. Nass, and O. John. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human-Computer Studies*, 66(5):303–317, 2008.

[4] J. Cacioppo and L. Tassinary. Inferring psychological significance from physiological signals. *American Psychologist*, 45:16–28, 1990.

[5] A. Choi and W. Woo. Physiological sensing and feature extraction for emotion recognition by exploiting acupuncture spots. *Lecture Notes in Computer Science*, 3784:590–597, 2005.

[6] J. A. Healey and R. W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportantion Systems*, 6:156–166, 2005.

[7] L. Heath. *Triangulation: Methodology*, pages 15901–15906. Elsevier Science Ltd.: Oxford, UK, 1 edition, 2001. ISBN: 978-0-08-043076-8.

[8] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis. Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, 38(3):502–512, 2008.

[9] K. H. Kim, S. W. Bang, and S. R. Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42:419–427, 2004.

[10] J. T. Larsen, G. G. Berntson, K. M. Poehlmann, T. A. Ito, and J. T. Cacioppo. *The psychophysiology of emotion*, chapter 11, pages 180–195. New York, NY, USA: Guilford, 3rd edition, 2008.

[11] C. L. Lisetti and F. Nasoz. Using noninvasive wearable computers to recogniza human emotions from physiological signals. *Journal of Applied Signal Processing*, 11:1672–1687, 2004.

[12] F. Nasoz, K. Alvarez, C. L. Lisetti, and N. Finkelstein. Emotion recognition from physiological signals for presence technologies. *International Journal of Cognition, Technology, and Work*, 6:4–14, 2003.

[13] E. Paulos. Connexus: A communal interface. In *Proceedings of the 2003 conference on Designing for user experiences*, pages 1–4, San Francisco, CA, USA, June 06–07 2003. New York, NY, USA: ACM.

[14] R. W. Picard, E. Vyzas, and J. A. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1175–1191, 2001.

[15] P. Rani, C. Liu, N. Sarkar, and E. Vanman. An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis & Applications*, 9(1):58–69, 2006.

[16] K. Takahashi. Remarks on emotion recognition from bio-potential signals. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume 2, pages 1655–1659, Palmerston North, New Zealand, 5–8 October 2003.

[17] N. Tractinsky. Tools over solutions? comments on interacting with computers special issue on affective computing. *Interacting with Computers*, 16(4):751–757, 2004.

[18] E. L. van den Broek, V. Lisý, J. H. Janssen, J. H. D. M. Westerink, M. H. Schut, and K. Tuinenbreijer. *Affective Man-Machine Interface: Unveiling human emotions through biosignals*, page [in press]. Berlin/Heidelberg, Germany: Springer, 2009.

[19] E. L. van den Broek, M. H. Schut, J. H. D. M. Westerink, and K. Tuinenbreijer. Unobtrusive Sensing of Emotions (USE). *Journal of Ambient Intelligence and Smart Environments*, 1(3):287–299, 2009.

[20] J. H. D. M. Westerink, M. Ouwerkerk, T. Overbeek, W. F. Pasveer, and B. de Ruyter. *Probing Experiences: From Academic Research to Commercial Propositions*, volume 8 of *Philips Research Book Series*. Springer: Dordrecht, The Netherlands, 2008.

[21] G. N. Yannakakis and J. Hallam. Entertainment modeling through physiology in physical play. *International Journal of Human-Computer Studies*, 66(10):741–755, 2008.

[22] J. Zhai and A. Barreto. Stress detection in computer users based on digital signal processing of noninvasive physiological variables. *Biomedical Science Instrumentation*, 42:495–500, 2006.