

Emotion Recognition Using Physiological Signals: Laboratory vs. Wearable Sensors

Martin Ragot^(✉), Nicolas Martin, Sonia Em, Nico Pallamin,
and Jean-Marc Diverrez

Usage and Acceptability Lab, b<>com, 1219, Avenue des Champs Blancs,
35510 Cesson-Sévigné, France
{martin.ragot,nicolas.martin,sonia.em,nico.pallamin,
jean-marc.diverrez}@b-com.com

Abstract. Emotion recognition is an important research topic. Physiological signals seem to be an appropriate way for emotion recognition and specific sensors are required to collect these data. Therefore, laboratory sensors are commonly used while the number of wearable devices including similar physiological sensors is growing up. Many studies have been completed to evaluate the signal quality obtained by these sensors but without focusing on their emotion recognition capabilities. In the current study, Machine Learning models were trained to compare the Biopac MP150 (laboratory sensor) and Empatica E4 (wearable sensor) in terms of emotion recognition accuracy. Results show similar accuracy between data collected using laboratory and wearable sensors. These results support the reliability of emotion recognition outside laboratory.

Keywords: Wearable sensors · Laboratory sensors · Emotion recognition · Machine learning · Physiological signals

1 Introduction

Emotion recognition is currently a hot topic in the field of affective computing [1]. In prior studies, several modalities have been explored to recognize the emotional states such as facial expression [2], speech [3], etc. However, the physiological signals related to autonomic nervous system appear as an appropriate way to assess objectively emotions [4]. Two types of sensors may be used for gathering physiological signals: laboratory and wearable sensors. Laboratory sensors seem effective [5] but, in some cases, they are not deployable outside controlled situations. Also, wearable sensors provide useful and non-obstructive way to obtain physiological signals [6, 7]. Moreover, the wearable sensors gathering physiological data become cheaper and widely available. The accuracy of these sensors has been explored in several studies and shows that the physiological signals gathered by laboratory sensors and wearable sensors seem quite similar. McCarthy and collaborators [6] indicate that the photoplethysmography (PPG) signals obtained from the Empatica E4 are sufficiently precise for the cardiac activity assessment. Other research [e.g., 7] validate wearable sensors as reliable and relevant for the physiological signals analysis. However, to the best of our knowledge, the emotion recognition accuracy obtained by different types of sensors [8]

was compared in only few studies. In the current study, laboratory and wearable sensors were used to gather physiological data with the aim to recognize emotional states using Machine Learning method (Support Vector Machine – SVM).

2 Method

2.1 Participants

We recruited 19 French volunteers via social networks: 12 women and 7 men whose average age was 33.89 years \pm 8.62 (see Table 1 for details). The minimum age is 23.49 years and maximum age is 52.46 years. Participants received €15 at the end of the experiment for their participation.

Table 1. Descriptive statistics of participants

	N	Mean age	SD age
Men	7	31.99	3.76
Women	12	35.00	10.51
Total	19	33.89	8.62

In the sample, all subjects were francophone. Participants had normal or corrected to normal vision. Moreover, participants had not taken any somatic drug(s), which may have an impact on physiological responses (e.g., corticosteroids), on the passing day.

To gather the physiological data, participants have been instrumented of two sensors: the Biopac MP150 (laboratory sensor) and Empatica E4 wristband (wearable sensor). Both sensors recorded cardiac and electrodermal (EDA) activities. In order to synchronize the two sensors during the data acquisition, a program was specifically developed in Python and C.

2.2 Material

For emotion induction, 45 color pictures extracted from the International Affective Picture System (IAPS) [9] have been displayed on a computer screen (1920 \times 1080 pixels). The valence and arousal associated to each picture were balanced. Thereby, each picture was categorized under three levels of valence (positive, neutral and negative) and three levels of arousal (high, medium and low) based on the theoretical values provided by the IAPS Technical Manual [9]. Finally, nine balanced categories were created (e.g., positive valence and low arousal) and five pictures of each category were presented to participants (the selected pictures ID are presented in Appendix 1) (Fig. 1).

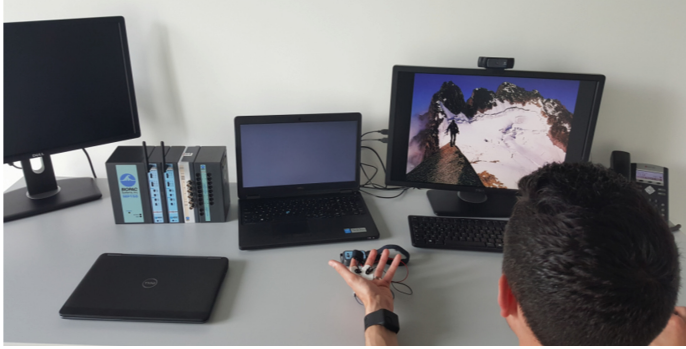


Fig. 1. Experiment setup

2.3 Physiological and Subjective Data

Subjective and physiological data have been collected during the experiment. Concerning the subjective data, two scales have been used. First, the Beck Depression Inventory II (BDI-II) [10] (21 items) was used before the experiment in order to exclude participants with depression issues. During the experiment, the Self-Assessment-Manikin (SAM) [11] was used to measure the emotional responses after each picture. Participant had to position himself on five different pictograms and four intermediate values (scoring from 1 to 9). As prior studies have shown that a 2-dimensional model of emotions (including valence and arousal) is preferable to a 3-dimensional model (including valence, arousal and dominance [12, 13]), only the evaluation of valence (positive/negative aspects of emotions) and arousal (emotional intensity) were considered.

Concerning the physiological data, EDA and cardiac activities have been recorded using two different sensors: the Biopac BioNomadix MP150 and Empatica E4 Wristband. Nine specific features have been extracted from these signals (HR, AVNN, SDNN, rMSSD, pNN50, LF, HF, RD and AVSCL). These features correspond to the most used features according to the literature review of Kreibitz [4].

2.4 Machine Learning

Machine Learning algorithms were used to consider the nonlinear relationship between subjective and physiological data. Machine learning models were trained in order to compare laboratory and wearable sensors in terms of emotion recognition accuracy. Support Vector Machine (SVM), supervised learning algorithms [14], were selected to classify data. After training, these models can recognize specific patterns related to specific outputs [15]. Technically, for the algorithms trainings, two types of data were used: physiological data as input and emotional states as output. After training, the models should be able to recognize the emotional states related to the physiological data.

To ensure genericity of the model, two main methods were used. First, the dataset was divided into training dataset (80%) and testing dataset (20%) (i.e., only the training dataset is used during the training). Second, cross-validation method was used during the training to improve the stability of results.

2.5 Procedure

At the beginning of the experiment, participants were informed of the experiment theme and signed a consent form. Then, participants completed a short general questionnaire (gender, date of birth, etc.). After, BDI-II was proposed to measure the clinical depression level (participants with a score ≥ 19 are excluded from the analyses). Afterwards, participants were instrumented with both sensors: the Empatica E4 wristband and Biopac MP150. In order to train participants to the subjective scale, a session with three pictures was also proposed (these data were excluded from the final analyses). Before each picture presentation, a black fixation cross on a white screen was displayed during 3 to 5 s (i.e., duration is randomly defined in order to limit expectation effect). The 45 pictures were presented randomly to participants while controlling the images sequence (i.e., two pictures from the same subcategory could not be displayed successively). Finally, after each picture presentation, participants had to evaluate their emotional state within 15 s using the Self-Assessment-Manikin (SAM) [11] through two dimensions: valence (positive/negative aspects of emotions) and arousal (emotional intensity).

3 Results

3.1 Descriptive Statistics

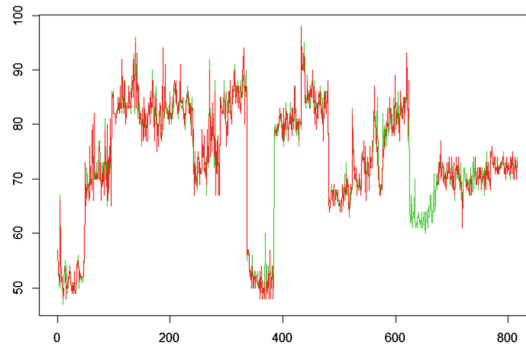
Among the subjective data collected, 2.3% of responses were missing (i.e., 20 missing subjective evaluations on the 871 collected). The average valence was $3.86 \pm .36$ where the score of 1 represents a very negative valence and 9 a very positive valence. The average arousal was 3.11 ± 1.55 where the score of 1 represents a very low arousal and 9 a very high arousal. The correlations were estimated between the features obtained from the Empatica E4 and Biopac MP150 data (see Table 2 for details). The correlations are high for the cardiac activity features (from .50 to .99). However, the correlation between AVSCL obtained by both sensors is low (.13)¹.

For illustration, the Heart Rate features extracted from both sensors are presented in Fig. 2.

¹ The weak correlation on EDA seems to be due to a problem of data recording for one participant. Deleting these data lead to a correlation of $r = .45$ between the AVSCL features gathered by both sensors.

Table 2. Correlations between the physiological features gathered by the Empatica E4 and Biopac MP150

	Biopac MP150									
Empatica E4		HR	AVNN	SDNN	rMSSD	pNN50	LF	HF	RF	AVSCL
	HR	.99								
	AVNN		.99							
	SDNN			.75						
	rMSSD				.69					
	pNN50					.61				
	LF						.57			
	HF							.58		
	RF								.50	
	AVSCL									.13

**Fig. 2.** HR signal gathered using the Empatica E4 (red line) and Biopac MP150 (green line). X-axis corresponding to the full dataset. Y-axis corresponding to the Heart Rate.**Table 3.** Results for emotion recognition

		Valence	Arousal
Empatica E4	Training	.657 (SD = .05)	.700 (SD = .02)
	Testing	.659	.704
Biopac MP150	Training	.655 (SD = .03)	.698 (SD = .05)
	Testing	.656	.697

In the “training” lines, two pieces of information are provided: the mean recognition rate through training sessions (cross-validation method) and number in brackets corresponding to standard deviation through training. The “testing” lines correspond to the recognition rate on the testing dataset.

3.2 Emotion Recognition System

Machine Learning algorithms (SVM) were used in order to consider the nonlinear relationships between these two types of data. The Machine Learning models were trained to recognize emotional states as binary variables. A training by sensor (i.e., Biopac M150 and Empatica E4) was carried out. For each sensor, two models were trained: one for valence and one for arousal. The Table 3 presents the main results. In summary, for both sensors, an accuracy of 66% for valence level and 70% for arousal level were found based on person-independent models.

According to these results, the accuracy of emotion recognition appears similar between the wearable sensor Empatica E4 and laboratory sensor Biopac MP150 in this experimental context.

4 Discussion and Conclusion

The aim of the current study was to compare the Empatica E4 and Biopac MP150 sensors in terms of emotion recognition capabilities. Thus, nine features were extracted from the physiological signals gathered by these sensors. The Machine Learning models were trained to recognize emotional states from these features. According to the results, the accuracy of emotion recognition appears similar with respectively an accuracy around 70% for arousal and 66% for valence.

In the current study, emotion recognition was based on extracted features. Consequently a strongly influence of these features on accuracy can be supposed. Nine features were used, a relatively weak number compared to some prior research [16]. Thus, extracting more features may lead to discover significant differences between sensors.

Overall, a stronger emotion induction may improve the accuracy of emotion recognition. Indeed, only few phasic responses have been detected (beyond the natural physiological responses) even though this feature reveals emotional activation. A stronger induction should influence physiological signals and may lead to difference between sensors. Thus, it could be interesting to conduct new studies to ensure of the similar recognition capabilities between sensors.

In future works, it could be relevant to compare emotion recognition from these sensors in a less controlled environment with the potential presence of motions.

In conclusion, in this study, wearable sensors appear as accurate as laboratory sensors for emotion recognition. The E4 device seems to be relevant for emotion recognition in daily life as a non-intrusive, easy to use and accurate wearable sensor.

Acknowledgments. We would like to thank all those who participated in any way in this research. This work was supported by the French government through the ANR Investment referenced ANR-10-AIRT-07.

Appendix 1

See Table 4.

Table 4. Picture ID by category of valence and arousal

Arousal	Positive valence	Neutral valence	Negative valence
High arousal	8492; 4659; 4695; 5629; 8501	1120; 5950; 8475; 1932; 8341	6300; 3301; 6263; 6520; 3500
Medium arousal	2075; 2160; 7330; 7470; 7580	8065; 7497; 2220; 1945; 5535	3550.1; 2345.1; 2800; 9140; 2751
Low arousal	5764; 5811; 1910; 5870; 2370	2101; 7038; 7185; 7490; 7491	9395; 2490; 9001; 2722; 2039

References

1. Petta, P., Pelachaud, C., Roddy, C. (eds.): Emotion-Oriented Systems: The Humaine Handbook. Springer, Berlin (2011)
2. Tian, Y., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 97–115 (2001)
3. Ververidis, D., Kotropoulos, C.: Emotional speech recognition: resources, features, and methods. *Speech Commun.* **48**, 1162–1181 (2006)
4. Kreibig, S.D.: Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* **84**, 394–421 (2010)
5. Jerritta, S., Murugappan, M., Nagarajan, R., Wan, K.: Physiological signals based human emotion recognition: a review. Presented at the IEEE 7th International Colloquium on Signal Processing and its Applications (CSPA), March 2011
6. McCarthy, C., Pradhan, N., Redpath, C., Adler, A.: Validation of the Empatica E4 wristband. Presented at the IEEE EMBS International Student Conference (ISC), May 2016
7. Garbarino, M., Lai, M., Tognetti, S., Picard, R., Bender, D.: Empatica E3 - a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. Presented at the EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth) (2014)
8. Ollander, S., Godin, C., Campagne, A., Charbonnier, S.: A comparison of wearable and stationary sensors for stress detection. Presented at the October 2016
9. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (IAPS): technical manual and affective ratings. The Center for Research in Psychophysiology, University of Florida, Gainesville, FL (1999)
10. Beck, A.T., Steer, R.A., Brown, G.K.: Beck Depression Inventory-II. Psychological Corporation, San Antonio (1996)
11. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **25**, 49–59 (1994)
12. Berlyne, D.E.: An Approach to Environmental Psychology, A. Mehrabian, J.A. Russell. MIT Press, Cambridge, MA (1974). *Behav. Ther.* **7**, 132–133 (1976)
13. Smith, C.A., Ellsworth, P.C.: Patterns of cognitive appraisal in emotion. *J. Pers. Soc. Psychol.* **48**, 813–838 (1985)

14. Kotsiantis, S.B.: Supervised machine learning: a review of classification techniques. In: Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, pp. 3–24. IOS Press, Amsterdam, The Netherlands (2007)
15. Bell, J.: Machine Learning: Hands-On for Developers and Technical Professionals. Wiley, Indianapolis (2014)
16. Janssen, J.H., van den Broek, E.L.: Guidelines for mobile emotion measurement. In: Geven, A., Tscheligi, M., and Noldus, L. (eds.) ACM MobileHCI Workshop? Measuring Mobile Emotions: Measuring the Impossible?, pp. 17–20. Center for Usability Research & Engineering (CURE), Vienna, Austria (2009)

Advances in Human Factors in Wearable Technologies
and Game Design

Proceedings of the AHFE 2017 International
Conference on Advances in Human Factors and
Wearable Technologies, July 17-21, 2017, The Westin
Bonaventure Hotel, Los Angeles, California, USA

Ahram, T.Z.; Falcão, C. (Eds.)

2018, XI, 268 p. 115 illus., Softcover

ISBN: 978-3-319-60638-5