A Comparative Study of Stress and Anxiety Estimation in Ecological Settings Using a Smart-shirt and a Smart-bracelet

Abhishek Tiwari[1], Raymundo Cassani[1], Shrikanth Narayanan[2], and Tiago H. Falk[1]
[1] Institut national de la recherche scientifique, INRS-EMT, Montreal, Canada
[2] University of Southern California, Los Angeles, USA

***Abstract*— In recent years, consumer wearable devices focused on health assessment have gained popularity. Of these devices, a large number target monitoring heart rate; a few among them include additional biometrics such as breathing rate, galvanic skin response, and skin temperature. Heart rate, and more specifically, heart rate variability (HRV) measures have proven useful in monitoring user psychological states, such as mental workload, stress and anxiety. Most studies, however, have been conducted in controlled laboratory environments with artificially-induced psychological responses. While these conditions assure high quality in the collected data, the amount of data are limited and the generalization of the findings to more ecologically-appropriate settings remains unknown. To this end, in this paper we compare the accuracy of two wearable devices, namely a smart-shirt measuring electrocardiograms and a smart-bracelet measuring photoplethysmograms. Several HRV features are extracted and tested as correlates of stress and anxiety. Data were collected from 196 participants during their normal work shifts for a period of 10 weeks. The complementarity of the two devices is also explored and the advantages of each method are discussed.**

I. Introduction

Advances in miniaturization, battery and sensing technologies have allowed the development of wearable devices for long-term, unobtrusive and continuous acquisition of biomedical data. With the emergence of the *quantified-self* movement [1], wireless heart rate monitors have proliferated not only within the clinical realm (e.g., [2]) but also within the sports and consumer markets (e.g., [3], [4]). A large number of these devices have reached the market in form of chest straps (e.g., Polar), smart-bracelets (Fitbit, Garmin), and smart-garments, which measure cardiac activity through sensors incorporated in the fabric (Hexoskin, OMsignal). Within these devices, heart rate (HR) has remained the principal measurement modality, as long-term monitoring of heart rate variability (HRV) has been shown useful for cardiovascular disease assessment [5], [6].

Additionally, metrics derived from HRV have been found to be important correlates of several quality-of-life indices, such as psycho-social workload [7] (i.e., job stressors), mental workload and anxiety [8], as well as mental fatigue [9]. Heart rate variability is an indicator of the changes in the autonomic nervous system and has traditionally been quantified using time- and/or frequency-domain features computed from inter-beat interval time series, also known as RR-intervals (RR_i). Time domain features quantify the statistical properties and rate of change of the RR series, whereas frequency domain features have been related to the balance between the sympathetic and parasympathetic branches of the nervous system [6], [10].

Typically, HRV studies have relied on controlled laboratory experiments with artificially induced psychophysiological responses, such as the use of Stroop tests for attention, or video games for stress. While this allows for high-quality data to be recorded and for subjective variability to be reduced, it limits the amount of data to be recorded, as well as duration of the data collection. Transferability of the obtained models and findings to more ecologically appropriate settings is also not assured. To overcome this limitation, we collected data from 196 staff members of a large hospital during a 10-week period. Participants wore a smart-shirt and a smart-bracelet, either alone or in combination, as they carried out their normal work shifts. With a companion App, they were asked about their day to day stress and anxiety levels. A comparison between the two devices is made in terms of the potential to measure stress and anxiety. Moreover, the complementarity of the two was also explored, suggesting that multimodal methods may be the most appropriate for "in-the-wild" experiments.

II. Materials and Methods

A. Participants

Data was collected from 196 participants (66 male, age 38.6 ± 9.8 years) from a pool of employees (nurses

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2017-17042800005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

978-1-5386-1311-5/19/$31.00 ©2019 Crown

and staff) of a large urban hospital in California. Approximately, two-thirds of the participants were nurses while one-third were hospital staff. During 10 weeks, participants carried out their work day as usual but were asked to fill a brief smartphone-based daily survey that included information on levels of stress and anxiety on a 5-point scale. Besides the daily surveys, during their work shift the participants wore multiple wearable sensors. Data collection protocol approval was obtained from the Institutional Review Board of the affiliated institutions and participants consented to participate in the study.

B. Wearable Sensors

Participants were outfitted with multiple wearable devices to collect a variety of biometric data such as vocal audio features [11], heart rate, respiration rate, and sleep quality, among others. Heart rate data were acquired simultaneously with an OMsignal smart-shirt and a Fitbit Charge 2 smart-bracelet. Henceforth, these devices are referred to as OMsignal and Fitbit, respectively. While HR measurement is reported by both devices, they are derived from different modalities. More specifically, OMsignal derives HR from the electrocardiogram (ECG) signal obtained with sensors embedded in the fabric of the shirt, while Fitbit extracts the HR from the photoplethysmogram (PPG) signal obtained at the wrist. Although both modalities are capable of obtaining reliable measurements of the instantaneous HR [12], only the OMsignal device provides HR values beat-to-beat, whereas the Fibit delivers a pre-processed coarse grained HR sample every 5 seconds. As such, the RR time series obtained from the HR values reported for each device have different temporal resolutions.

In addition to HR, the devices provide other biometric and activity measurements. For example, OMsignal monitors breathing rate, step count and intensity of physical activity. The device also provides an internal quality parameter called RR coverage (RR_{cov}), which is the number of R peaks clearly detected over a 5-minute period, with higher values indicating higher ECG quality of signal. Fitbit, in turn, provides step count, sleep time and quality, and calories burned. For this study, only the heart rate data is used for comparisons. A summary of the functionalities of each device and the modalities measured is presented in Table I.

C. HRV Feature Extraction

Here, classical time- and frequency-domain HRV features were extracted from the RR time series. A complete description of these HRV measurements can be found in [6], [10]. Time-domain features were computed for both devices, but due to the low temporal resolution of Fitbit, not all frequency-domain features could be measured for the smart-bracelet. Time- and frequency-domain features were extracted over 5-minute windows. A complete list of the extracted features is given in Table II along with an indication if they were extracted for OMsignal, Fitbit, or both. It is important to emphasize that the majority of these features have been shown in the literature to correlate with mental workload [13] and anxiety [8].

TABLE I
FUNCTIONALITY AND MODALITY COMPARISON BETWEEN THE TWO WEARABLE DEVICES USED.

	OMSignal	Fitbit Charge 2
Technique	ECG	PPG
HR resolution	Instantaneous	Avg every 5 s
Form factor	Shirt or Bra	Bracelet
Battery life	18-20 hrs	5 days
Other measurements	breathing rate, physical activity	sleep, calories, steps

TABLE II
HRV FEATURES COMPUTED FROM THE RR TIME SERIES.

	Feature	OMsignal	Fitbit
Time	mean RR	x	x
	standard deviation (std) RR (SDNN)	x	x
	coefficient of variation RR	x	x
	% of RR differences > 50 ms (pNN50)	x	x
	mean of 1^{st} RR difference (diff)	x	x
	std of absolute (abs) 1^{st} RR diff	x	x
	root mean square of 1^{st} RR diff (RMSDD)	x	x
	mean abs 1^{st} RR diff of normalized RR	x	x
Frequency	total power	x	x
	high frequency power (HF)	x	
	ratio of HF to total power (HF norm)	x	
	low frequency power (LF)	x	
	ratio of LF to total power. (LF norm)	x	
	ratio of LF to HF (LF/HF)	x	
	very low frequency power (VLF)	x	

The HRV features were further aggregated over an entire day using the following statistical functionals: mean, standard deviation, coefficient of variation, median, min, max, 1st and 3rd quartile, skewness, and kurtosis. Thus, for a day of data from each participant, a total of 90 features derived from Fitbit and 150 features derived from OMSignal were computed. The prediction power for the stress and anxiety levels was evaluated for three features sets, namely, (*i*) OMsignal, (*ii*) Fitbit, and (*iii*) combined OMsignal-Fitbit. The latter explores the complementary nature of the two devices and the usefulness of a multimodal setup.

D. Feature Selection and Classification

As we are looking at changes in cardiac activity over the duration of a day, despite the long duration of the data collection protocol, we are still left with a relatively small number of samples relative to the number of features explored. As such, feature selection needs to be performed in order to not only sift out the most discriminatory features, but also to remove features that are highly correlated. For this purpose, recursive feature elimination was performed with a step size of 5 using the Extra Trees Classifier. The top 50 features were then selected for classification at each cross-validation step. Feature selection and classification were performed on OMsignal, Fitbit and OMsignal-Fitbit feature sets. A five-fold cross-validation test setup was performed with feature selection taking place for the top 50 features at each fold. Classification was then performed on subject-wise binarized high/low stress and anxiety levels. A support vector machine (SVM) classifier with a RBF kernel (radius=1/number of features) was used. The 'balanced' setting in the SVM classifier, where the target value is used to automatically adjust class weights of the inversely proportional to class frequencies in the input data. These class weights are then used to readjust the penalty parameter C to $C*(class_weight[i])$ for class i. Finally, as the data was unbalanced, F1-score, balanced accuracy (BACC), sensitivity (Sens), and specificity (Spec) are used as classifier performance figures-of-merit. Our experiments were carried out with the open-source scikit-learn toolbox [14].

III. Experimental Results and Discussion

A. Cross-Device Similarity

Pearson's correlation between the nine common HRV features from OMsignal and Fitbit was computed to gauge the similarity in HR reading between the two modalities. Table III reports these correlation values for all segments (corresponding to the $RR_{cov} \geq 0$), as well as to identify the segments deemed high quality ($RR_{cov} \geq 0.8$) by the smart-shirt. As can be seen, both devices capture heart rate accurately, with a correlation of 0.84 when noisy OMsignal segments are considered and 0.92 if only high-quality segments are used. Interestingly, the correlations between HRV features, despite their increase with increasing RR_{cov}, are much lower and mostly below 0.5. This low correlation is due to the lower temporal resolution of Fitbit, thus potentially limiting its usage for long-term HRV analysis.

B. Stress and Anxiety Predictions

Classification results for stress and anxiety are shown in Tables IV and V, respectively. As can be seen, for stress measurement OMsignal-based features outperformed Fitbit features in all figures-of-merit used (e.g., 4.12% higher BACC, and 3.14% higher F1 score). Combining both modalities only resulted in slight improvements in BACC, F1, and sensitivity. For anxiety, OMSignal features again performed better than the Fitbit ones across all tested figures-of-merit (e.g., 4.19% higher BACC and 5.95% higher F1 score). Unlike stress measurement, however, for anxiety monitoring the multimodal OMsignal-Fitbit feature set resulted in improvements across all tested performance metrics, with only a slight drop in specificity (0.6331 versus 0.6335). Overall, gains over OMSignal features alone of 1.49% BACC and 2.13% F1-score could be seen. For anxiety measurement, Fitbit features alone achieved very low sensitivity, but helped improve overall sensitivity when combined with OMsignal features, thus suggesting their complementarity. These findings corroborate those of [15], [16] and suggest that multimodal approaches may be useful for monitoring of stress and anxiety in natural real-world settings such as work place, as extraction of HRV features from varying temporal resolutions seems

TABLE III
CORRELATION BETWEEN OMSIGNAL AND FIBIT FEATURES FOR DIFFERENT RR COVERAGE VALUES

Feature	$RR_{cov} \geq 0$	$RR_{cov} \geq 0.8$
mean RR	0.84	0.92
SDNN	0.39	0.53
coefficient of variation RR	0.27	0.46
RMSDD	0.19	0.33
pNN50	0.21	0.35
mean of 1^{st} RR diff	0.06	0.35
std of abs 1^{st} diff	0.13	0.21
mean of abs 1^{st} diff norm RR	0.15	0.29
total power	0.07	0.19

TABLE IV
PERFORMANCE COMPARISON FOR STRESS PREDICTION

Feature	BACC	F1	Sens	Spec
OMSignal	0.5847	0.5481	0.5389	0.6305
Fitbit	0.5435	0.5167	0.5261	0.5610
Combined	0.5869	0.5546	0.5517	0.6222

TABLE V
PERFORMANCE COMPARISON FOR ANXIETY PREDICTION

Feature	BACC	F1	Sens	Spec
OMSignal	0.5765	0.4906	0.5195	0.6335
Fitbit	0.5346	0.4311	0.4419	0.6273
Combined	0.5914	0.5119	0.5497	0.6331

to provide complementary information, likely due to the fractal nature of cardiac signals.

To further explore the importance of the features, an in-depth analysis on the features ranked highly across all five cross-validation trials was performed on the fused data set. For stress, we observed that 14 features were common in 4 of the 5 feature selection steps while 15 features were common in all the 5 steps. Of the total 29 features, 8 features were from Fitbit and included different functional aggregates of mean of RR, coefficient of variation, and mean of absolute 1^{st} difference of normalized RR. Of the Fitbit features, the coefficient of variation was most commonly occurring with 5 different functional aggregates. From OMSignal, in turn, most frequency features appeared in the top 29 features, except for total power in all frequencies and high frequency power. Additionally the coefficient of variation and mean RR each occurred 4 times with different functional aggregates.

For anxiety, we observed that 17 features were common in 4 of the 5 feature selection steps although no features appeared in the top fifty feature set for all 5 steps. Of the 17 features only 2 features were from Fitbit and consisted in different functional aggregates of the coefficient of variation of the RR series. The most common feature from OMSignal was the mean of absolute 1^{st} difference of the normalized RR series, which appeared 6 times, with different functional aggregates. From the frequency features only very low frequency power appeared in the 17 features twice as median and 1^{st} quartile functional aggregates. Overall, we found that skewness is the most commonly occurring functional aggregate in the top features for stress, while max and coefficient of variation are commonly occurring aggregates for anxiety.

IV. Conclusions

In this work we compared stress and anxiety prediction performance based on heart rate variability features computed from a smart-shirt (OMsignal) and a smart-bracelet (Fitbit). It was observed that despite using different monitoring modalities (ECG versus PPG), both devices measured HR similarly, achieving correlations as high as 0.92. Their varying temporal resolutions, however, resulted in varying HRV parameters. Overall, HRV features computed from OMsignal data (higher temporal resolution) proved to be more effective at predicting stress and anxiety. Notwithstanding, by combining features from both OMsignal and Fitbit, it was found that improvements could be obtained, particularly for anxiety prediction, thus suggesting that using features at varying temporal resolutions or scales may be useful to characterize the fractal behavior of cardiac activity.

References

[1] D. Nafus and J. Sherman, "Big data, big questions| this one does not go up to 11: the quantified self movement as an alternative big data practice," *International journal of communication*, vol. 8, p. 11, 2014.

[2] R. Fensli, T. Gundersen, T. Snaprud, and O. Hejlesen, "Clinical evaluation of a wireless ecg sensor system for arrhythmia diagnostic purposes," *Medical engineering & physics*, vol. 35, no. 6, pp. 697–703, 2013.

[3] H. J. Wellens and A. P. Gorgels, "How important is the electrocardiogram in protecting and guiding the athlete?" 2011.

[4] D. J. Plews, P. B. Laursen, J. Stanley, A. E. Kilding, and M. Buchheit, "Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring," *Sports medicine*, vol. 43, no. 9, pp. 773–781, 2013.

[5] P. Melillo *et al.*, "Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis," *PloS one*, vol. 10, no. 3, p. e0118504, 2015.

[6] A. Camm *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[7] F. Togo and M. Takahashi, "Heart rate variability in occupational health-a systematic review," *Industrial health*, vol. 47, no. 6, pp. 589–602, 2009.

[8] W.-H. Wen *et al.*, "Toward constructing a real-time social anxiety evaluation system: Exploring effective heart rate features," *IEEE Transactions on Affective Computing*, 2018.

[9] M. Patel, S. K. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert systems with Applications*, vol. 38, no. 6, pp. 7235–7242, 2011.

[10] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Frontiers in Public Health*, vol. 5, Sep. 2017.

[11] T. Feng, A. Nadarajan, C. Vaz, B. Booth, and S. Narayanan, "Tiles audio recorder: an unobtrusive wearable solution to track audio activity," in *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*. ACM, 2018, pp. 33–38.

[12] D. T. Weiler, S. O. Villajuan, L. Edkins, S. Cleary, and J. J. Saleem, "Wearable Heart Rate Monitor Technology Accuracy in Research: A Comparative Study Between PPG and ECG Technology," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1, pp. 1292–1296, Sep. 2017.

[13] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, and L. Pecchia, "Acute mental stress assessment via short term hrv analysis in healthy adults: A systematic review with meta-analysis," *Biomedical Signal Processing and Control*, vol. 18, pp. 370–377, 2015.

[14] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[15] A. L. Goldberger, L. A. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley, "Fractal dynamics in physiology: alterations with disease and aging," *Proceedings of the national academy of sciences*, vol. 99, no. suppl 1, pp. 2466–2472, 2002.

[16] M. D. Costa, C.-K. Peng, and A. L. Goldberger, "Multiscale analysis of heart rate dynamics: entropy and time irreversibility measures," *Cardiovascular Engineering*, vol. 8, no. 2, pp. 88–93, 2008.