Measuring Behavior 2020-21

Volume 1

12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods

13-15 October, 2021, Kraków, Poland

Proceedings

Volume Editors

Andrew Spink

Noldus Information Technology; Andrew.Spink@noldus.nl

Jaroslaw Barski

Medical University of Silesia, Katowice; jbarski@sum.edu.pl

Anne-Marie Brouwer

Perceptual and Cognitive Systems, TNO; anne-marie.brouwer@tno.nl

Gernot Riedel

University of Aberdeen; g.riedel@abdn.ac.uk

Annesha Sil

University of Aberdeen; annesha.sil@abdn.ac.uk

ISBN: 978-90-74821-93-3

DOI: 10.6084/m9.figshare.13013717

Validation of wearables for electrodermal activity (EdaMove) and heart rate (Wahoo Tickr)

Ana Borovac [1], Ivo Stuldreher [2], Nattapong Thammasan [3] and Anne-Marie Brouwer [2]

1 Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavík, Iceland. anb48@hi.is.

2 Perceptual and Cognitive Systems, TNO, Soesterberg, The Netherlands. ivo.stuldreher@tno.nl. anne-marie.brouwer@tno.nl

3 Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands. n.thammasan@utwente.nl

We are interested in monitoring physiology out of the lab because of its potential informative value about mental state, such as stress. For real life studies or applications utilizing physiology, we need to equip individuals with wearables that are designed to be compatible with individuals performing their daily activities, which can come at the cost of a reduction in signal quality. In the current study, we compare wearable sensors of electrodermal activity (EdaMove) and heart rate (Wahoo Tickr) to their high-end, laboratory counterparts (ActiveTwo). Signals were compared at a general level as well as in relation to their response to emotional sounds. EdaMove showed more general drift and responses returned slower to baseline than ActiveTwo. Responses to emotional sounds were about equally clear as ActiveTwo. Apart from a delay of around 6.7 seconds, Wahoo Ticker accurately recorded general heart rate levels. However, it did not capture fast changes which also resulted in less clear responses to emotional sounds. Both wearables are potentially suitable to record physiology in real life, but for synchrony recordings, EdaMove is expected to be suitable whereas results based on Wahoo Tickr are expected to be less clear.

Introduction

Monitoring physiological measures such as heart rate and skin conductance (or electrodermal activity - EDA) in daily life can be of interest for measuring human behavior. Monitoring a physiological variable can be helpful for tracking certain medical condition (such as cardiac activity for people with a cardiac disease) and also for gaining insights about a person's mental state, such as stress, which has a major influence on behavior and health. While relating physiology to mental state is not without challenges [1], in a recent systematic review, De Looff et al. [2] report that high levels of job stress are associated with an increased heart rate, and decreased heart rate variability measures. For EDA, they reported that such evidence was not available - no EDA studies were found that met the inclusion criteria of the review.

Monitoring physiology outside the lab requires wearable devices. For heart rate, many types of (affordable) wearable devices are available. For EDA this is not the case, which may partly explain the scarcity of real-life studies involving EDA. However, given that EDA is uniquely innervated by the sympathetic *fight or flight* part of the Autonomic Nervous System, it may offer valuable information related to stress. Indeed, recent studies show the first evidence that EDA as recorded during daily life can be related to individuals' burnout symptoms over time [3] and EDA increases preceding aggressive behavior [4]. Sano et al. [5] showed EDA features to be informative about students' level of stress and mental health. These three studies all used a particular EDA wearable that records EDA with sensors embedded in a wristband. While this is convenient for the user, EDA responses to arousing events as recorded using sensors at the wrist have been found to be much less sensitive compared to sensors at the palm of the hand or fingertips, the latter being the usual location to record EDA in the laboratory [6,7]. Specifically, Van Lier et al. [8] show in a thorough study on validating wearables that a wristband could pick up effects caused by a strong sustained stressor, but, in contrast to the *gold standard* sensors placed at the finger, not to short, more mild ones. These findings probably reflect a lower density of sweat glands at the wrist compared to the fingers or palm [9].

Volume 1 of the Proceedings of the joint 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods to be held in Krakow, Poland, October 15-18, 2021. www.measuringbehavior.org.
Editors: Andrew Spink, Jarosław Barski, Anne-Marie Brouwer, Gernot Riedel, & Annesha Sil (2020).
ISBN 978-90-74821-93-3. DOI: 10.6084/m9.figshare.13013717

We are interested in recording physiological responses to events that may be arousing for an individual in real life. In the current study, we tested heart rate and EDA wearables that we estimated to be relatively sensitive to pick up such responses. Therefore, we did not use wearables from the side of the spectrum that may have favored user comfort at the cost of signal quality (i.e., we did not use wristbands for recording EDA and heart rate) but wearables that are less comfortable to use on a daily basis, but rely on higher quality signals. These wearables are Wahoo Tickr, a wearable that extracts heart rate from electrical signals as measured at the chest, and EdaMove, a wearable that measures skin conductance at the hand palm. While wristbands are easier to attach, and in the case of EDA, less visible and interfering than a chestband and sensors at the palm, we found both of these wearables to be working well from a user perspective in recordings in which 90 young teenagers were recorded during a school day.

We collected EDA and HR in an experiment aimed at examining the effect of attentional instruction in multiple physiological data streams. It involved auditory emotional stimuli, and EDA and HR were recorded using both high-end laboratory equipment as well as the EdaMove and Wahoo Tickr. We utilize the data of this experiment here to compare signals coming from both types of equipment at a general level as well as at a response level, i.e. in relation to an emotional sound.

Methods

Participants

We recorded from 27 participants (aged between 18 and 48) with no self-reported problems in hearing or attention. Prior to the experiment all participants signed informed consent and after the experiment they received a small monetary award for their time and traveling costs. Data of three participants were discarded due to failed physiological recordings. The study was approved by TNO Institutional Review Board (TCPE) and TU Delft Human Research Ethics Committee.

Materials

Standard EDA, ECG (electrocardiogram) and EEG were recorded using an ActiveTwo system (BioSemi, Amsterdam, Netherlands) at 1024 Hz. EEG data is not further discussed here. Additionally, a wearable device, EdaMove 4 (Movisens GmbH, Karlsruhe, Germany), was also used to recorded EDA at different positions at 32 Hz. Figure 1 shows a picture of the simultaneous recording of EDA using both lab-grade and consumer-grade equipment with recording locations at the fingers and the palm respectively. HR was also recorded using Wahoo Tickr (Wahoo Fitness, Atlanta, GA, USA).

For EDA recorded with ActiveTwo, two passive gelled Nihon Kohden electrodes were placed on the ventral side of the distal phalanges of the middle and index finger. For EDA recorded with EdaMove, two self-adhesive electrodes were placed on the palm. Electrodes of both devices were placed on the non-dominant hand.

For ECG recorded with ActiveTwo, two active gelled Ag-AgCl electrodes were placed at the right clavicle and lowest floating left rib. For recordings with Wahoo Tickr, a band was fitted around the participant’s chest after applying gel on its sensors.

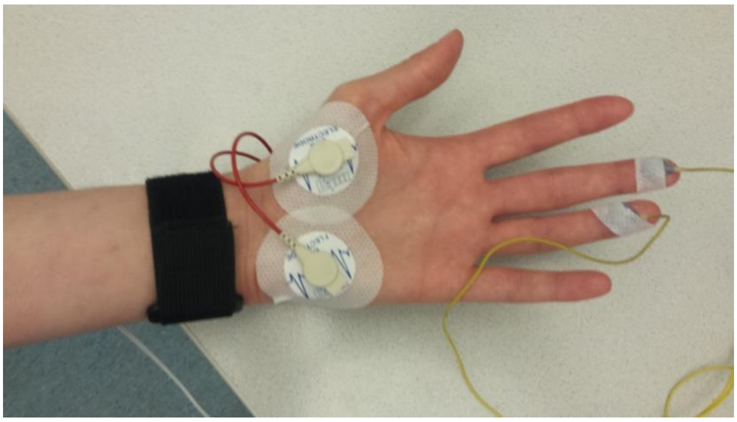

Figure 1. Electrodes on two fingers record EDA with ActiveTwo and electrodes on palm record EDA with EdaMove.

Stimuli and design

Each participant listened to an audio file, composed of a 66 min audiobook interspersed with other auditory stimuli including emotional sounds. Emotional sounds were taken from the IADS (International Affective Digitized Sounds-[10]: 12 neutral, 12 pleasant and 12 unpleasant sounds. Half of the participants were asked to attend to the audiobook and half to the other stimuli, where we expect the emotional stimuli to draw the attention of all participants.

Analysis

Data processing was done using MATLAB 2018b software (The Mathworks, Natick, MA, USA). ActiveTwo EDA was first downsampled to 32 Hz. The phasic (fast) and tonic (slow) components were extracted for both ActiveTwo and EdaMove EDA signals using Continuous Decomposition Analysis [11] as implemented in the Ledalab toolbox for MATLAB (available on www.ledalab.de).

The output from Wahoo Tickr is HR at 1 Hz. ActiveTwo ECG was downsampled to 256 Hz and high-pass filtered at 0.5 Hz. R-peaks were detected from a squared version of the reconstructed frequency-localized version of the ECG waveform using wavelets, following The Mathworks (2015). The R-to-R interval, or inter-beat interval (IBI) was extracted. The IBI semi-time series was transformed into a time-series HR. This was done by interpolating consecutive IBIs and then resampling at 2 Hz.

For both EDA and HR, plots of the raw wearable and ActiveTwo signals across the whole experiment were inspected for all participants.

To examine their sensitivity to emotional stimuli, EDA and HR signals were also aligned to the onset of the emotional sounds. For EDA, the phasic component was used since here, we are interested in the fast changes. In addition, the number of phasic EDA peaks larger than 0.1 μS was determined for both signals from the 1st to the 65th minute of the experiment.

Results

EDA

Figure 2 shows a typical sample of raw EDA traces from both devices. We observed that the return to baseline after a peak is slower in EDA recorded with EdaMove than with ActiveTwo. Figure 3 shows this in averaged data.

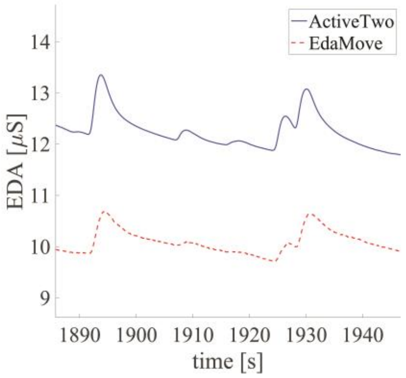

Figure 2. Example of raw EDA traces.

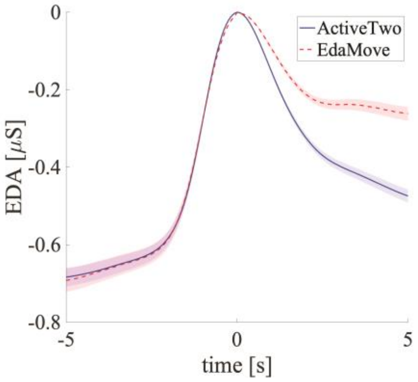

Figure 3. Averages of peaks with the standard error of the mean.

In this figure, epochs of raw signals were aligned to the time and the height of the maxima of the peaks. In order to include the same peaks present in both EdaMove and ActiveTwo signals, only peaks present in phasic components of both devices were used. A peak was considered to be present in both phasic components if the time difference of its maximum value was not larger than 1 s. Total number of peaks varied strongly between participants (a range of 0-570 for EdaMove and 1-358 for ActiveTwo). A paired *t*-test indicated no significant difference (t_{23}=1.85, *p*=0.08).

Besides the difference in the response shape, visual inspection also revealed more slow drift in the signal of the EdaMove than in ActiveTwo. An impression of this is given by the averages of the tonic components of all participants in Figure 4.

Figure 5 shows phasic EDA baselined on the onset of the emotional stimulus that elicited the strongest EDA response in our experiment (EroticFem2).

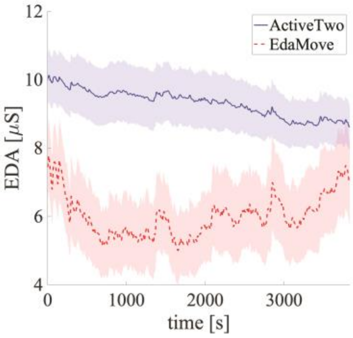

Figure 4. Averages of the tonic component of EDA signal with the standard error of the mean.

Volume 1 of the Proceedings of the joint 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods to be held in Krakow, Poland, October 15-18, 2021. www.measuringbehavior.org.
Editors: Andrew Spink, Jarosław Barski, Anne-Marie Brouwer, Gernot Riedel, & Annesha Sil (2020).
ISBN 978-90-74821-93-3. DOI: 10.6084/m9.figshare.13013717

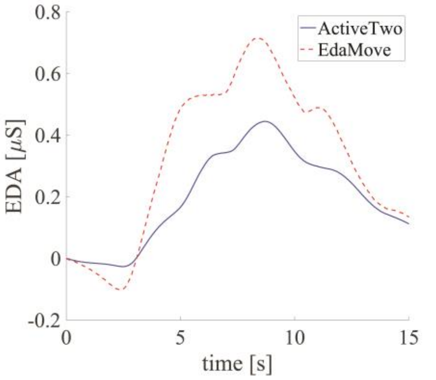

Figure 5. Averages of the phasic component of EDA signal after the IADS EroticFem2 stimulus.

Inspecting these plots for all sounds shows that averaged EDA phasic responses to emotional sounds are similar for both devices. If anything, EdaMove averaged phasic responses are somewhat stronger than those recorded by ActiveTwo. This may be caused by the slower drop of the signal as described above, causing a strong average response even when response latencies between individuals differ.

HR

Figure 6 shows a typical sample of HR traces from both devices. ActiveTwo reflects the typical, fluctuating pattern of HR that is associated with breathing [12]. HR from Wahoo Tickr does not show this. Since HR as obtained from Wahoo Tickr was delayed with respect to that obtained from ActiveTwo, Figure 6 also shows a shifted version of the Wahoo Tickr data. Shifting data was done using the determined delay of that participant. For this we filtered Wahoo Tickr data with a moving average (window size of 20) and filtered ActiveTwo HR with a 4th order low pass digital Butterworth filter with a normalized cutoff frequency of 0.03. After filtering, minima and maxima with a prominence larger than 1 were detected. We then calculated the time difference between peaks in Wahoo Tickr and ActiveTwo signals. Extrema are counted as the same if the delay of the Wahoo Ticker is not larger than 10 s. Finally, the median of the found delays of an individual participant was defined as the final delay of the Wahoo Tickr. The mean delay across participants was 6.7 s (sd: 1.3).

The mean absolute HR difference was 3.4 bpm (median: 2.7, sd: 2.2, range: 1.9-11.1). Mean Wahoo Tickr HR was 69.5 bpm and ActiveTwo HR was 70.1 bpm. This was not significantly different (paired *t*-test: t_{23}=1.89, p=0.07).

Figure 7 shows HR baselined on the onset of EroticFem2. A decrease in HR in response to the sound can be observed for both devices (consistent with Bradley & Lang [13]). Consistent with our finding that Wahoo Tickr does not capture fast fluctuating changes in HR, this pattern is more pronounced for ActiveTwo.

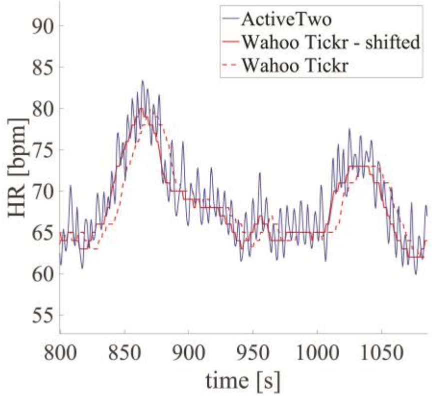

Figure 6. Example of raw HR traces. The dashed line indicates the Wahoo Tickr HR before shifting for the calculated delay.

Volume 1 of the Proceedings of the joint 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods to be held in Krakow, Poland, October 15-18, 2021. www.measuringbehavior.org.
Editors: Andrew Spink, Jarosław Barski, Anne-Marie Brouwer, Gernot Riedel, & Annesha Sil (2020).
ISBN 978-90-74821-93-3. DOI: 10.6084/m9.figshare.13013717

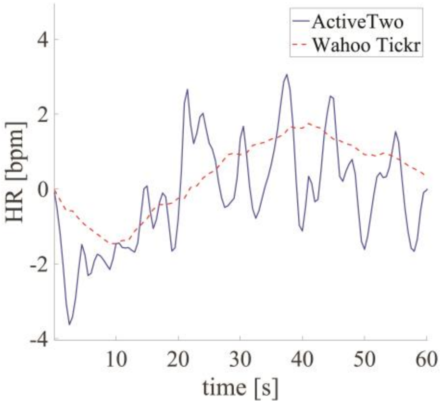

Figure 7. HR with the standard error of the mean after the IADS EroticFem2 stimulus.

Discussion

We observed some systematic differences between the EDA and HR signals from the wearable and lab-grade equipment.

Return to baseline seemed to be slower for EDA signals recorded with EdaMove than ActiveTwo. The underlying reason is yet to be investigated but we speculate that the size of EdaMove's self-adhesive electrodes affect the return part of a peak by slowing down evaporation of sweat. This might also induce more drift in the signal. The number of peaks found in phasic component of EDA did not significantly differ, though EdaMove tends to show more peaks. We observed for some parts of the data that EdaMove showed peaks that were completely absent for ActiveTwo, suggesting that this has something to do with the distribution of sweat glands at the lower end of the palm versus the fingers rather than with sensitivity of the sensors.

HR from Wahoo Tickr was delayed with respect to ActiveTwo and did not capture fast changes. This is likely caused by filtering and other processing by the Wahoo Tickr equipment. The difference in absolute HR between the systems are caused by the fact that only ActiveTwo captured fast changes in HR such as those caused by breathing, but also by ActiveTwo capturing ectopic beats (‘skipping’ a beat, or an ‘extra’ beat) that are usually excluded from the data for a more robust estimate of HR. This happened relatively often in three participants.

We found both wearables to perform quite well. Slow drifts (EdaMove) or a delay (Wahoo Tickr) are not necessarily big problems for analyses of emotional responses as long as they are known. General HR as captured by Wahoo Tickr corresponds to the ActiveTwo gold standard. EdaMove captured quick changes as those caused by emotional sounds equally well as ActiveTwo, while this was more difficult Wahoo Tickr. When the aim is to capture physiological responses to potentially emotional events, we therefore expect EdaMove to perform about as well as ActiveTwo and Wahoo Tickr as less well. In addition, the lack of systematic fast variation in HR (Figure 6) indicates that Wahoo Tickr is not suited to capture heart rate variability.

Note that in the present study, we recorded from participants who were sitting still. In real life circumstances involving movement, recordings of EDA and HR will be affected by noise and inherent associations of EDA and HR with body movement. This is a topic of ongoing research. See e.g. [14] for a study of EdaMove under conditions of movement, and a way to identify movement artefacts, and [15] for potential ways ot cope with the effect of movement on HR.

References

1. Brouwer A.-M., Zander T.O., van Erp J.B.F., Korteling J.E., Bronkhorst A.W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in Neuroscience* **9**:136.

2. de Looff P.C., Cornet L.J.M., Embregts P.J.C.M., Nijman H.L.I., Didden H.C.M. (2018). Associations of sympathetic and parasympathetic activity in job stress and burnout: A systematic review. *PLOS ONE* **13**(**10**):e0205741.

3. de Looff P., Didden R., Embregts P., Nijman H. (2019). Burnout symptoms in forensic mental health nurses: Results from a longitudinal study. *International Journal of Mental Health Nursing* **28(1)**:306-317.

4. de Looff P., Noordzij M.L., Moerbeek M., Nijman H., Didden R., Embregts P. (2019). Changes in heart rate and skin conductance in the 30 min preceding aggressive behavior. *Psychophysiology* **56(10)**: e13420.

5. Sano A., Taylor S., McHill A.W., Phillips A.J., Barger L.K., Klerman E., Picard R. (2018). Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study. *Journal of Medical Internet Research* **20(6)**: e210.

6. Payne A.F.H., Schell A.M., Dawson M.E. (20182016 Lapses in skin conductance responding across anatomical sites: Comparison of fingers, feet, forehead, and wrist. *Psychophysiology* **53(7)**:1084-1092.

7. van Dooren M., de Vries J.J.G., Janssen J.H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology and Behavior* **106(2)**: 298-304.

8. van Lier H.G., Pieterse M.E., Garde A., Postel M.G., de Haan H.A., Vollenbroek-Hutten M.M.R., Schraagen J.M., Noordzij M.L. (2019). A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the e4 biosensor. *Behavior Research Methods*, 1-23.

9. Boucsein W. (2012). *Electrodermal activity*. New York: Springer Science+Business Media, LLC.

10. Bradley M.M., Lang P.J. (2007). The international affective digitized sounds (IADS-2): Affective ratings of sounds and instruction manual. *Technical Report B-3*, University of Florida , Gainesville, FL.

11. Benedek M., Kaernbach C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods* **190(1)**: 80-91.

12. Aasman J., Mulder G., Mulder L.J. (1987). Operator effort and the measurement of heart-rate variability. *Human factors* **29(2)**: 161-170.

13. Bradley M.M., Lang P.J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology* **37(2):** 204-215.

14. Thammasan N., Stuldreher I.V., Wismeijer D., Poel M., van Erp J.B.F., Brouwer A.-M. (2019). A novel, simple and objective method to detect movement artefacts in electrodermal activity. *Proceedings of 8th International Conference on Affective Computing and Intelligent Interaction* (Cambridge, 3-6 September 2019), 1-7.

15. Brouwer A.-M., van Dam E., van Erp J.B.F., Spangler D.P., Brooks J.R. (2018). Improving real-life estimates of emotion based on heart rate: A perspective on taking metabolic heart rate into account. *Frontiers in Human Neuroscience* **12**, 284.