# A Comparison between Laboratory and Wearable Sensors in the Context of Physiological Synchrony

Jasper J. van Beers
Perceptual and Cognitive Systems
TNO
Soesterberg, The Netherlands
jasper.vanbeers@tno.nl

Ivo V. Stuldreher†
Perceptual and Cognitive Systems
TNO
Soesterberg, The Netherlands
ivo.stuldreher@tno.nl

Nattapong Thammasan
Human Media Interaction
University of Twente
Enschede, The Netherlands
n.thammasan@utwente.nl

Anne-Marie Brouwer
Perceptual and Cognitive Systems
TNO
Soesterberg, The Netherlands
anne-marie.brouwer@tno.nl

## ABSTRACT

Measuring concurrent changes in autonomic physiological responses aggregated across individuals (Physiological Synchrony - PS) can provide insight into group-level cognitive or emotional processes. Utilizing cheap and easy-to-use wearable sensors to measure physiology rather than their high-end laboratory counterparts is desirable. Since it is currently ambiguous how different signal properties (arising from different types of measuring equipment) influence the detection of PS associated with mental processes, it is unclear whether, or to what extent, PS based on data from wearables compares to that from their laboratory equivalents. Existing literature has investigated PS using both types of equipment, but none compared them directly. In this study, we measure PS in electrodermal activity (EDA) and inter-beat interval (IBI, inverse of heart rate) of participants who listened to the same audio stream but were either instructed to attend to the presented narrative (n=13) or to the interspersed auditory events (n=13). Both laboratory and wearable sensors were used (ActiveTwo electrocardiogram (ECG) and EDA; Wahoo Tickr and EdaMove4). A participant's attentional condition was classified based on which attentional group they shared greater synchrony with. For both types of sensors, we found classification accuracies of 73% or higher in both EDA and IBI. We found no significant difference in classification accuracies between the laboratory and wearable sensors. These findings encourage the use of wearables for PS based research and for in-the-field measurements.

†Corresponding author: ivo.stuldreher@tno.nl (I. V. Stuldreher)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
*ICMI'20, October 25-29, 2020, Virtual event, Netherlands.*
© 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-7581-8/20/10...$15.00.
DOI: https://doi.org/10.1145/3382507.3418837

## CCS CONCEPTS

• Applied computing → Law, social and behavioral sciences

## KEYWORDS

Selective attention; Physiological synchrony; wearable sensors; autonomic physiology; auditory; group

**ACM Reference format:**

Jasper J. van Beers, Ivo V. Stuldreher, Nattapong Thammasan, and Anne-Marie Brouwer, 2020. A Comparison between Laboratory and Wearable Sensors in the Context of Physiological Synchrony. In *Proceedings of 2020 ACM International Conference on Multimodal Interaction (ICMI'20), October 25-29, Virtual event, Netherlands.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3382507.3418837

## 1 Introduction

Autonomic physiological responses can provide informative insights into an individual's cognitive and emotional state. When aggregated across multiple individuals, group level dynamics may be investigated through similarities in their physiological activity. This concept is known as physiological synchrony (PS).

One prevalent domain in which autonomic PS has been extensively employed is that of interpersonal interactions [1]. Use of PS can also be envisioned in various human-computer interactions, such as in competitive multiplayer video games. Interest in autonomic PS is gaining traction in part due to the rapid development and growing maturity of wearable sensor technology [2]-[4] compounded with the possibility to combine a myriad of different wearable devices. Moreover, synchrony within different physiological modalities could be reflective of different processes, such as stress or empathy [1], and may provide insight into the mechanisms driving PS. It is suspected that PS can be used to measure shared attention [5], where shared attention may be an underlying explaining factor of other findings such as those by [6] on audience engagement. Attention itself plays an important role in learning capabilities [7], [8], task performance [9], and social interactions [10]. Jamet et al. [8]

demonstrated the benefits of using attention-guiding techniques to facilitate learning, resulting in improved performance for retention (e.g. memory) based tasks. As such, PS may be of interest as a tool to monitor attention continuously and unobtrusively in the classroom to assist students with learning disabilities, or to improve upon existing teaching methods [11].

Wearable sensors are typically unobtrusive, affordable, and mobile, enabling 'in-the-field' research which may provide for more realistic insights into natural human behavior. However, these benefits usually come at the cost of a diminished signal. For instance, the Wahoo Tickr, a wearable used to measure heart rate (HR), high frequency HR information is lost due to the low sampling rate and on-board processing [12] (In Press). It is still unclear what physiological signal aspects are relevant for measuring PS and how the limitations imposed by wearables influence their ability to measure meaningful affective/cognitive PS. Therefore, we chose to directly compare PS obtained through both laboratory and wearable sensors. Indeed, there are a few studies which compare wearable and laboratory grade equipment, such as that conducted by Ragot et al. [13] on recognizing emotion. However, these lie outside the domain of PS and thus may rely on different signal aspects.

To directly compare laboratory and wearable equipment, data obtained from an experiment described in [5] and [14] were used. In this experiment, participants were instructed to listen to the same audio track, but to attend to different stimulus aspects with the aim of determining the selective attention in groups using PS. In [5] and [14], EDA and HR data obtained from laboratory equipment (ActiveTwo) were analyzed. During the experiment, EDA and HR were concurrently measured with wearable sensors (Wahoo Tickr, HR; EdaMove4, EDA). The current study elaborates on this experiment and directly compares autonomic PS results between wearable data and their laboratory counterparts, when subject to the same conditions and analysis methods. Therein, we aim to evaluate the feasibility of the use of wearables, spanning two physiological modalities, in the domain of PS. To the best of our knowledge, this is the first study that compares PS from wearable data with PS from high-end laboratory equipment.

# 2 Methods

## 2.1 Participants

Participants (N = 27, aged between 18 and 48), with no self-reported problems in hearing or attention, were recruited from the research institute's (TNO) participant pool. All participants signed an informed consent form prior to the experiment and were given a small monetary reward after the experiment. Data of one participant was removed due to failed recordings. The study was approved by the TNO Institutional Review Board (TCPE) and the TU Delft Human Research Ethics Committee.

## 2.2 Materials

For the laboratory equipment, both EDA and electrocardiogram (ECG) were measured via an ActiveTwo system (BioSemi, Amsterdam, Netherlands) at 1024 Hz. For EDA, two passive gelled Nihon Kohden electrodes were placed on the ventral side of the distal phalanges of the middle and index finger on participants' left hand. For ECG, two active gelled Ag-AgCl electrodes were placed at the right clavicle and lowest floating left rib. Regarding the wearable equipment, EDA was recorded through an EdaMove4 (movisens GmbH, Karlsruhe, Germany) at 32Hz while HR was measured with a Wahoo Tickr (Wahoo Fitness, Atlanta, Georgia, USA) at 1Hz. The EdaMove4 was attached by two self-adhesive electrodes placed on the palm on participants' left hand. The Wahoo Ticker was fitted around the chest of participants after applying gel on its sensors. The Wahoo Tickr outputs a filtered HR signal with a minimum increment of 1 bpm derived from a measured electrical signal and thus does not provide raw inter-beat intervals (IBI). The signal resolution (1 bpm) and sampling rate (1 Hz) are independent of each other, but the amount of information contained in the signal is dependent on both. Thus, the Wahoo Tickr lacks high frequency HR information (e.g. respiratory sinus arrhythmia - RSA), that are present in the ActiveTwo ECG.

## 2.3 Stimuli and Design

Each participant completed the experiment individually and listened to the same audio file. This file was composed of a 66 min audiobook (a Dutch thriller written by Corine Hartman: 'Zure koekjes') with interspersed auditory stimuli (beeps and affective sounds). These short stimuli were randomly ordered with intervals between stimuli ranging from 35 to 55 seconds. Half of the participants were assigned to attend to the narrative (NA) of the audiobook and to ignore all other stimuli. The other half of the participants were asked to focus on the short stimuli (SSA) and ignore the narrative.

The affective sounds, of 6 second durations, are taken from the International Affective Digitized Sounds (IADS) [15]: a collection of acoustic stimuli normatively rated for emotion, valence and dominance. Examples include sounds of a crying baby or the cheers of a sports crowd. 12 neutral sounds, 12 pleasant sounds and 12 unpleasant sounds were elected. Beeps were presented in blocks of 30 seconds, with every two seconds a 100ms high (1kHz) or low (250Hz) pitched beep. SSA participants were tasked with counting the number of high and the number of low tones [16]. 27 blocks of sounds were presented.

## 2.4 Analysis

Data processing was done using MATLAB R2018b (Mathworks, Natick, MA, USA).

ActiveTwo EDA measurements were downsampled to 32 Hz. For both ActiveTwo and EdaMove4, the phasic component of the EDA response was extracted for further analysis using the Ledalab toolbox for MATLAB [17]. Studies on EDA typically show a certain number of 'non-responders' [18], or weak responders - participants with a low EDA magnitude and near-zero phasic response. Weak responders in our study were identified through visual inspection by the individual authors of the manuscript. Data of these participants were not discarded since the weak responses of these participants appeared to contain information pertaining to the shape of the response, which may be useful for synchrony. However, the phasic responses of the EdaMove4 weak responders were contaminated with peaks arising from noise and jitter due to on-board processing, distorting the signal shape. Therefore, the full EDA traces of the EdaMove4 weak responders were filtered using a Savitzky-Golay filter with a three second window and the phasic components were recomputed. Since the experiment event markers are expressed in the ActiveTwo timeline, we accounted for delays between EdaMove4 and ActiveTwo, among others arising from on-board processing on EdaMove4. The phasic response obtained via EdaMove4 was time corrected through a normalized cross-correlation with the phasic response from ActiveTwo. Here, the lag maximizing the correlation of the two signals is assumed to represent the accumulated delay (inconsistent across participants).

ECG measurements acquired from ActiveTwo were first downsampled to 256Hz, then high-pass filtered at 0.5Hz. R-peaks of the ECG signal were detected following [19]. The resulting semi-timeseries of consecutive IBIs were subsequently interpolated and resampled at 32Hz to transform them into a timeseries. The Wahoo Tickr HR signal was first upsampled to 32 Hz, then time corrected through a normalized cross-correlation with the ActiveTwo derived HR (i.e. inverse of IBI). The pre-processed Wahoo Tickr HR was then converted to IBI.

Regardless of sensor type and physiological signal, inter-subject correlations (ISC) were determined using a moving window, as introduced by [20]. A window of size 15 seconds traverses the signal at 1 second increments with Pearson correlations calculated over successive windows. The overall correlation between two responses is given by the natural logarithm of the sum of all positive correlations divided by the absolute value of the sum of all negative correlations. Classifications were based on the average ISC of a participant with all members from the NA group and all members from SSA group, excluding the participant in question. Participants were classified with the attentional group that they were more correlated with (i.e. shared the highest ISC). Paired sample t-tests were conducted to determine whether the NA ISC and SSA ISC were significantly different within each attentional group for EDA and IBI. Chance level classifications were determined through surrogate data with 100 instances of randomly shuffled attentional group labels. To evaluate if the classification accuracies between the laboratory and wearable sensors are statistically different, an exact McNemar's test was used. This test is suitable to compare paired nominal data with small sample sizes, such as ours [21], [22].

## 3 Results

Figure 1 illustrates that the patterns in ISC are similar between the laboratory and wearable sensors, and that overall, within-group ISC is higher than between-group ISC for both types of equipment. Figure 1 also presents results for both NA and SSA participants. For EDA, the within-group ISC is significantly higher than the between-group ISC for NA participants with the EdaMove4 ($t(12) = 4.16, p = .001$) but not for the ActiveTwo ($t(12) = 0.74, p = .476$). For SSA participants, within-group ISC is significantly higher than between-group ISC for the ActiveTwo ($t(12) = 4.07, p = .002$) but not for the EdaMove4 ($t(12) = 1.98, p = .072$). Regarding IBI, the significance of the patterns in group-level ISC are consistent between the ActiveTwo and Wahoo Tickr. For SSA participants, the within-group ISC is significantly higher than the between-group ISC (ActiveTwo: $t(12) = 2.27, p = .043$ ; Wahoo Tickr: $t(12) = 4.75, p < .001$). Within-group ISC is not significantly higher than between-group ISC for NA participants (ActiveTwo: $t(12) = 2.02, p = .066$; Wahoo Tickr: $t(12) = 1.38, p = .192$).

Table 1 presents the percentage of participants whose attentional condition (i.e. NA or SSA) was correctly identified. The corresponding chance level classification accuracies are also shown in accompanying brackets. For EDA, the classification accuracy is significantly above chance for both ActiveTwo ($t(99) = 2.39, p = .009$) and EdaMove4 ($t(99) = 2.91, p = .005$). Likewise, the IBI classification accuracy is significantly above chance for both ActiveTwo ($t(99) = 2.43, p = .009$) and Wahoo Tickr ($t(99) = 3.38, p = .001$). Rather than performing worse, wearables tend to outperform their laboratory counterparts. EdaMove4 classifies 81% of the participants correctly as opposed to 73% with ActiveTwo. Similarly, Wahoo

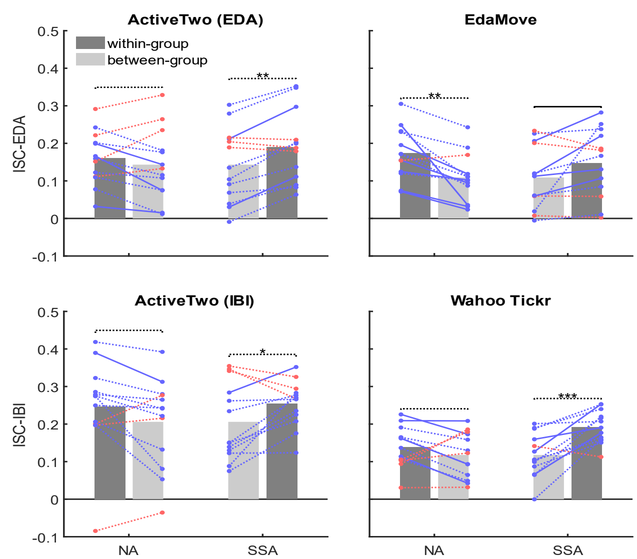


**Figure 1: Within-group and between-group inter-subject correlations (ISC) of electrodermal activity (EDA, top) and inter-beat interval (IBI, bottom) for both attentional groups (NA, left bars; SSA, right bars) derived from laboratory (ActiveTwo, left column) and wearable (EdaMove and Wahoo Tickr, right column) sensors. Also illustrated are connected dots which represent the individual participants. Full blue lines indicate higher within-group ISC. Dotted red lines denote higher between-group ISC. Paired sample t-tests were used to determine if within-group ISC are significantly higher than between-group ISC ($* p < .05, ** p < .01, *** p < .001$).**

**Table 1: Overall classification accuracy (in percentage) of correctly identified participant attentional conditions based on their inter-subject correlations (ISC). The chance level values, and associated standard deviations, are given in brackets. The corresponding *p*-values are also presented.**

| | ActiveTwo | Wearables (EdaMove4; Wahoo Tickr) |
|---|---|---|
| EDA | 73 (50 $\pm$ 10) | 81 (52 $\pm$ 9) |
| | $p = .009$ | $p = .005$ |
| IBI | 77 (50 $\pm$ 11) | 81 (50 $\pm$ 9) |
| | $p = .009$ | $p = .001$ |

Tickr classifies 81% of the participants correctly, in comparison to 77% with ActiveTwo. However, an exact McNemar's test showed no statistical difference between the classification accuracy of ActiveTwo and EdaMove4, $p = .7266$, or between ActiveTwo and Wahoo Tickr, $p = 1.000$.

## 4 Discussion

Through this study, we have shown that PS in selective attention can be derived from wearable sensors, EdaMove4 and Wahoo Tickr, equally well as their laboratory-based counterparts.

Our results are especially notable for the Wahoo Tickr, given its poor resolution (1 bpm) and sampling rate (1Hz). This suggests that wearables with lower bitrates may also be appropriate for PS-based research, broadening the potential applications of autonomic PS. The relatively good performance of the Wahoo Tickr could suggest that the very-low to low frequency HR (i.e. 0.003 to 0.15Hz) [23] is an influential feature for determining synchrony in selective attention. The lower frequency components of the ActiveTwo and Wahoo Tickr HR traces are mostly coincident, hence, the presence of high frequency HR (e.g. due to breathing) could act as 'noise' and may explain some of the differences in classification performance between these sensors. Consequently, future work should investigate methods to remove breathing from the ActiveTwo data to compare results more directly with the Wahoo Tickr. Under conditions of movement, synchrony in HR due to shared attention may be strengthened, if the movements are associated with shared attention such as in [23], or overshadowed when unrelated as seen in [24]. However, any synchrony in HR induced by quick breathing pattens, as with [23], will not be captured by the wearable sensor used here.

Differences between EdaMove4 and ActiveTwo can, in part, be explained through the 'weak responders'. In total, there were three weak responders for ActiveTwo, two of which were in the SSA group, and seven weak responders for EdaMove4, five of which were in the SSA group. The large concentration of weak responders among the EdaMove4 SSA participants may explain the difference in significance of the group-level ISC between ActiveTwo and EdaMove4 seen at the top of Figure 1. Phasic responses of the EdaMove4 weak responders may have lacked some synchrony relevant features (e.g. peaks) which were either filtered out or were not present in the initial signal, resulting in poorer classification performance. For instance, two participants who were weak responders for EdaMove4 but not for ActiveTwo were misclassified with EdaMove4 data and correctly classified with ActiveTwo data. This misclassification of weak responders is not unique to EdaMove4 since ActiveTwo also misclassifies some weak responders. In general, weak responders are difficult to classify due to a lack of informative features. This lack of features may also artificially suppress the magnitude of the group-level ISC, leading to unreliable classifications which extend beyond these weak responders. To mitigate this, a different physiological modality (such as IBI) may be used to compliment the classification result. In the current study, all but one of the EdaMove4 weak responders were correctly identified by the Wahoo Tickr, motivating the use of various modalities to augment classification accuracies.

For participants who were not weak responders, any discrepancies in performance between ActiveTwo and EdaMove4 may be explained by the long recovery time of the EdaMove4 [12]. We suspect that the long recovery time is due to the large adhesive pads of EdaMove4 which impair the evaporation of sweat. In this region, the magnitude of the phasic response is locally reduced while the noise level remains constant. This culminates in a lower signal-to-noise ratio within the affected region and mirrors challenges observed with weak responders. Moreover, this locally reduced response may artificially inflate synchrony since this feature has a large temporal footprint and is present across all participants (high chance of overlap between participants). Due to this, using EdaMove4 may be limited to experiments which aim to measure synchrony across temporally sparse events, or that combine various physiological modalities.

The exact McNemar's tests show that there is no statistical difference in classification accuracy between wearables and laboratory equipment when measuring PS in shared attention, and the trend is even such that wearables perform better rather than worse. Clearly, our findings encourage the use of wearables for PS based experiments and for in-the-field research. Limitations of this study are that the experiment was conducted in laboratory conditions with minimal movement and that only two types of wearables were used for comparison. Therefore, it is yet unclear as to how appropriate other wearables are for computing PS and how suitable wearables in general for more active applications.

## 5 Conclusion

The current study indicates that measuring PS in shared attention with laboratory and wearable sensors can result in similar performance between the two. PS derived from the wearable sensors used in this study distinguished between the two attentional conditions (NA and SSA) equally well as PS obtained from laboratory equipment, in both physiological modalities (EDA and IBI). Since wearables are less obtrusive and are inherently mobile, these results motivate the use of wearable sensors for both in-the-lab and in-the-field measurements, such as for measuring PS in an audience during artistic performances or students in a classroom.

# REFERENCES


[1] Richard V. Palumbo, Marisa E. Marraccini, Lisa L. Weyandt, Oliver Wilder-Smith, Heather A. McGee, Siwei Liu and Matthew S. Goodwin, 2017. Interpersonal Autonomic Physiology: A Systematic Review of the Literature, *Personality and Social Pscyhology Review* 21, 2, 99-141. DOI: https://doi.org/10.1177/1088868316628405.

[2] Subhas C. Mukhopadhyay, 2015. Wearable Sensors for Human Activity Monitoring: A Review, *IEEE Sensors Journal* 15, 3, 1321-1330. DOI: https://doi.org/10.1109/JSEN.2014.2370945.

[3] Dhruv R. Seshadri, Ryan T. Li, James E. Voos, Janes R. Rowbottom, Celeste M. Alfes, Christian A. Zorman and Colin K. Durmmond, 2019. Wearable sensors for monitoring the physiological and biochemical profile of the athlete, *npj Digital Medicine* 2. DOI: https://doi.org/10.1038/s41746-019-0150-9.

[4] Hendrika G. van Lier, Marcel E. Pieterse, Ainara Garde, Marloes G. Postel, Hein A. de Haan, Miriam M. R. Vollenbroek-Hutten, Jan M. Schraagen and Matthijs L. Noordzij, 2019. A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor, *Behavior Research Methods* 52, 607-629. DOI: https://doi.org/10.3758/s13428-019-01263-9.

[5] Anne-Marie Brouwer, Ivo V. Stuldreher and Nattapong Thammasan, 2019. Shared attention reflected in eeg, electrodermal activity and heart rate, *2019 Proceedings of the Workshop Socio-Affective Technologies: an interdisciplinary approach co-located with IEEE SMC 2019 (Systems, Man and Cybernetics)*, 27-31. Retrieved from: http://ceur-ws.org/Vol-2474/shortpaper6.pdf.

[6] Shkurta Gashi, Elena Di Lascio and Silvia Santini, 2019. Using unobtrusive wearable sensors to measure the physiological synchrony between presenters and audience members. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1, DOI: https://doi.org/10.1145/3314400.

[7] Yuhong Jiang and Marvin M. Chun, 2001. Selective attention modulates implicit learning, *The Quarterly Journal of Experimental Psychology: Section A* 54, 4, 1105-1124. DOI: https://doi.org/10.1080/02724980042000516.

[8] Eric Jamet, Monica Gavota and Christophe Quaireau, 2008. Attention guiding in multimedia learning, *Learning and Instruction* 18, 2, 135-145. DOI: https://doi.org/10.1016/j.learninstruc.2007.01.011.

[9] Harold Pashler, James C. Johnston and Eric Ruthruff, 2001. Attention and Performance, *Annual Review of Psychology* 52, 629-651. DOI: https://doi.org/10.1146/annurev.psych.52.1.629.

[10] Brendan F. Andrade, Darlene A. Brodeur, Daniel A. Waschbusch, Sherry H. Stewart and Robin McGee, 2009. Selective and Sustained Attention as Predictors of Social Problems in Children With Typical and Disordered Attention Abilities, *Journal of Attention Disorders* 12, 4, 341-352. DOI: https://doi.org/10.1177/1087054708320440.

[11] Suzanne Dikker, Lu Wan, Ido Davidesco, Lisa Kaggen, Matthias Oostrik, James McClintock, Jess Rowland, Georgios Michalareas, Jay J. Van Bavel, Mingzhou Ding, and David Poeppel, 2017. Brain-to-Brain syncrhony tracks real-world dynamic group interactions in the classroom, *Current Biology* 27, 9, 1375-1380. DOI: https://doi.org/10.1016/j.cub.2017.04.002.

[12] Ana Borovac, Ivo V. Stuldreher, Nattapong Thammasan and Anne-Marie Brouwer, 2020. Validation of wearables for electrodermal activity (EdaMove) and heart rate (Wahoo Tickr), *Proceedings of Measuring Behaviour 2020-21*, (In Press).

[13] Martin Ragot, Nicolas Martin, Sonia Em, Nico Pallamin and Jean-Marc Diverrez, 2017. Emotion Recognition Using Physiological Signals: Laboratory vs. Wearable Sensors, *Advances in Human Factors and Eronomics* 608, 15-22. DOI: https://doi.org/10.1007/978-3-319-60639-2_2.

[14] Ivo V. Stuldreher, Nattapong Thammasan, Jan B. F. van Erp and Anne-Marie Brouwer, 2020. Physiological synchrony in EEG, electrodermal activity and heart rate reflects shared selective auditory attention, *Journal of Neural Engineering* 17, 046028. DOI: https://doi.org/10.1088/1741-2552/aba87d.

[15] Margaret M. Bradley and Peter J. Lang, 2007. The International Affective Digitized Sounds (IADS-2): Affective ratings of sounds and instruction manual, University of Florida, Gainesville FL.

[16] Alix L. de Dieuleveult, Anne-Marie Brouwer, Petra C. Siemonsma, Jan B. F. van Erp and Eli Brenner, 2018. Aging and sensitivity to illusory target motion with or without secondary tasks, *Multisensory Research* 31, 3-4, 227-249. DOI: https://doi.org/10.1163/22134808-00002596.

[17] Mathias Benedek and Christian Kaernbach, 2010. A continuous measure of phasic electrodermal activity, *Journal of Neuroscience Methods* 190, 1, 80-91. DOI: https://doi.org/10.1016/j.jneumeth.2010.04.028.

[18] Jason J. Braithwaite, Derrick G. Watson, Robert Jones, Mickey Rowe, 2013. A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments, *Psychophysiology* 49, 1, 1017-1034. Retrieved from: https://www.birmingham.ac.uk/Documents/college-les/psych/saal/guide-electrodermal-activity.pdf

[19] Jiapu Pan and Willis J. Tompkins, 1985. A Real-Time QRS Detection Algorithm, *IEEE Transactions on Biomedical Engineering* 32, 3, 230-236. DOI: https://doi.org/10.1109/TBME.1985.325532.

[20] Carl D. Marci, Jacob Ham, Erin Moran and Scott P. O. Orr, 2007. Physiologic Correlate of Perceived Therapist Empathy and Social-Emotional Process During Psychotherapy, *The Journal of Nervous and Mental Disease* 195, 2, 103-111. DOI: https://doi.org/10.1097/01.nmd.0000253731.71025.fc.

[21] Omolola A. Adedokun and Wilella D. Burgess, 2012. Analysis of Paired Dichotomous Data: A Gentle Introduction to the McNemar Test in SPSS, *Journal of MultiDisciplinary Evaluation* 8, 17, 125-131. ISSN: 1556-8180.

[22] Giles M. Foody, 2009. Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority, *Remote Sensing of Environment* 113, 8, 1658-1663. DOI: https://doi.org/10.1016/j.rse.2009.03.014.

[23] Giorgio Quer, Joshal Daftari and Ramesh R. Rao, 2016. Heart Rate Wavelet Coherence Analysis to Investigate Group Entrainment, *Pervasive and Mobile Computing* 28, C, 21-34. DOI: https://doi.org/10.1016/j.pmcj.2015.09.008.

[24] Kevin J. Verdière, Mélisande Albert, Frédéric Dehais, and Raphaëlle N. Roy, 2020. Physiological Synchrony Revelaed by Delayed Coincidence Count: Application to a Cooperative Complex Environment, *IEEE Transactions in Human-Machine Systems.* DOI: https://doi.org/10.1109/THMS.2020.2986417.