



Physiological Synchrony in EEG, Electrodermal Activity and Heart Rate Detects Attentionally Relevant Events in Time

Ivo V. Stuldreher^{1,2*}, Nattapong Thammasan², Jan B. F. van Erp^{1,2} and Anne-Marie Brouwer¹

¹ Perceptual and Cognitive Systems, Netherlands Organisation for Applied Scientific Research (TNO), Soesterberg, Netherlands, ² Human Media Interaction, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, Netherlands

OPEN ACCESS

Edited by:

Hasan Ayaz,
Drexel University, United States

Reviewed by:

Lucas C. Parra,
City College of New York (CUNY),
United States
Motoaki Kawanabe,
Advanced Telecommunications
Research Institute International (ATR),
Japan

*Correspondence:

Ivo V. Stuldreher
ivo.stuldreher@tno.nl

Specialty section:

This article was submitted to
Neural Technology,
a section of the journal
Frontiers in Neuroscience

Received: 23 June 2020

Accepted: 10 November 2020

Published: 03 December 2020

Citation:

Stuldreher IV, Thammasan N,
van Erp JBF and Brouwer A-M (2020)
Physiological Synchrony in EEG,
Electrodermal Activity and Heart Rate
Detects Attentionally Relevant Events
in Time. *Front. Neurosci.* 14:575521.
doi: 10.3389/fnins.2020.575521

Interpersonal physiological synchrony (PS), or the similarity of physiological signals between individuals over time, may be used to detect attentionally engaging moments in time. We here investigated whether PS in the electroencephalogram (EEG), electrodermal activity (EDA), heart rate and a multimodal metric signals the occurrence of attentionally relevant events in time in two groups of participants. Both groups were presented with the same auditory stimulus, but were instructed to attend either to the narrative of an audiobook (audiobook-attending: AA group) or to interspersed emotional sounds and beeps (stimulus-attending: SA group). We hypothesized that emotional sounds could be detected in both groups as they are expected to draw attention involuntarily, in a bottom-up fashion. Indeed, we found this to be the case for PS in EDA or the multimodal metric. Beeps, that are expected to be only relevant due to specific “top-down” attentional instructions, could indeed only be detected using PS among SA participants, for EDA, EEG and the multimodal metric. We further hypothesized that moments in the audiobook accompanied by high PS in either EEG, EDA, heart rate or the multimodal metric for AA participants would be rated as more engaging by an independent group of participants compared to moments corresponding to low PS. This hypothesis was not supported. Our results show that PS can support the detection of attentionally engaging events over time. Currently, the relation between PS and engagement is only established for well-defined, interspersed stimuli, whereas the relation between PS and a more abstract self-reported metric of engagement over time has not been established. As the relation between PS and engagement is dependent on event type and physiological measure, we suggest to choose a measure matching with the stimulus of interest. When the stimulus type is unknown, a multimodal metric is most robust.

Keywords: physiological synchrony, inter-subject correlations, interpersonal physiology, EEG, electrodermal activity, heart rate, multimodal

INTRODUCTION

Knowing what events in the external environment people attend to, and how their shared attentional engagement to events varies over time, can be useful in a range of settings, from evaluating educational or entertaining material, to real time adjustment of important instructions. Unlike explicit measures, such as questionnaires in which people are asked to specify their attentional engagement, physiological signals can provide continuous and implicit information on mental state (Zander and Kothe, 2011). However, the link between mental state and physiological measures [e.g., electroencephalography (EEG), electrodermal activity (EDA) or heart rate] is not straightforward (Brouwer et al., 2015). A popular approach to uncover the complex links between physiology and mental state is the use of supervised learning algorithms. These algorithms predict mental state based on a set of features extracted from physiological variables (Hamadicharef et al., 2009; Hussain et al., 2011; Fleureau et al., 2012; Liu et al., 2013; Aliakbaryhosseinabadi et al., 2017). A disadvantage of these types of analyses is the need for labeled training data, i.e., a set of physiological data that are labeled with a known value for the mental state of interest. Not only is it time consuming to obtain such a labeled dataset, it is also very difficult to determine the 'ground truth' mental state than can be used for data labeling (Brouwer et al., 2015). A second drawback of these supervised learning approaches is that classification is often limited to a small number of discrete states. Attentional, emotional or cognitive state, however, cannot realistically be represented by a small number of discrete states, but are naturally of more continuous nature (Zehetleitner et al., 2012; Rosenberg et al., 2013).

For monitoring attentional engagement, an approach that may be suited to circumvent both of the abovementioned problems is to monitor the physiological synchrony (PS) between individuals. PS is the degree to which physiological measures of multiple people uniformly change. Studies exploring PS in functional magnetic resonance imaging data have revealed strong voxel-wise inter-subject correlations across participants exposed to a common narrative stimulus (Hasson et al., 2004, 2010; Hanson et al., 2009). In the faster EEG signals, similar results were found (Dmochowski et al., 2012, 2014). The fast-changing EEG enabled the computation of a continuous measure of PS in time and suggested that moments of high PS corresponded with emotionally arousing scenes of the movie clips (Poulsen et al., 2017). For instance, high PS was found when scenes were viewed that involved the threat of a gun. Dmochowski et al. (2014) further showed that moment-to-moment variation in the PS predicted the general expressions of interest and attention of the public as indicated by number of tweets during a popular television series. Davidesco et al. (2019) found that PS over time indicated what specific information was retained by students in a lecture. Namely, PS was higher in lecture parts that provided answers for questions that students answered incorrectly in the pre-test and correctly in the delayed post-test than for questions where students' answers did not change. The relationship between neural PS and attentional engagement was also found to be less complex than most traditional physiological metrics.

Neural PS was found to be directly proportional to attentional engagement, as strong correlations were found between PS and performance on questionnaires reflective of paid attention (Cohen and Parra, 2016; Cohen et al., 2018; Stuldreher et al., 2020). This directly proportional relationship may thus be used to circumvent supervised learning approaches and the problems that come with such approaches, such as the dependency on labeled training data.

In the current work, we aim to employ the relation between PS and attentional engagement to detect the occurrence of attentionally relevant events in time. Rather than limiting the analyses to EEG, we also include PS measures of peripheral nervous system activity (EDA and heart rate), and quantify their comparative sensitivity of detecting relevant events. Up to recently, PS in peripheral physiological measures has been studied mainly as a metric of some form of affective connectedness between individuals (reviewed by Palumbo et al., 2017). Examples include peripheral PS in therapist-patient dyads as a measure of psychotherapy success (Koole et al., 2020), in couples in marital counseling as a measure of therapy outcome (Tourunen et al., 2019) and as measure of collaborative learning (Malmberg et al., 2019). Positive results found in these contexts may (partly) be driven by shared attentional engagement to external events, as connectedness between people may be strongly associated with mutual attentiveness (Tickle-Degnen and Rosenthal, 1990). Recently, it was found that PS in EDA and heart rate can indeed reflect shared attention toward narrative stimuli (Pérez et al., 2020; Stuldreher et al., 2020).

The advantage of peripheral physiological measures over EEG is that they can be recorded more easily and less obtrusively. In addition, EEG and peripheral measures may complement each other since they likely reflect different mental processes. EEG is, for example, sensitive to selective attention (Polich, 2007), whereas EDA and heart rate are sensitive to (emotional) arousal (Cacioppo et al., 2007; Boucsein, 2012).

As of yet, it is unknown whether PS in EEG, EDA and heart rate can be used to detect relevant moments in time. For EDA and heart rate, time-resolved dynamics of PS have not been investigated at all in the context of attentional engagement. For EEG, time-resolved dynamics have been explored (see for instance Dmochowski et al., 2014; Poulsen et al., 2017), but this has not been done systematically, using a-priori known cognitively or emotionally engaging stimuli for which detection performance can be evaluated. We here evaluate whether PS in EEG, EDA and heart rate can be used to detect cognitively or emotionally relevant moments in time. Our goal is not to compare detection performance directly between the different types of stimuli, but to evaluate PS for a range of events differing in terms of total duration, sound onset, mental processes addressed and more. Just as in real-world conditions, some event may capture attention in a bottom-up fashion, related to salience or emotional relevance, whereas others may only capture attention due to top-down mechanisms related to task instruction (Lang, 1995; Öhman et al., 2001; Schupp et al., 2003). We invited participants to come to our lab and listen to an audiobook that was interspersed with short auditory events, that we expected to induce emotional and cognitive load. We divided

the participants in two equal-sized groups. Participants in the audiobook-attending group (AA) were instructed to focus their attention on the audiobook and ignore the interspersed stimuli. Participants in the stimulus-attending group (SA) were instructed to focus their attention on the interspersed stimuli and ignore the audiobook. In a previous paper on this experiment (Stuldreher et al., 2020), we showed that PS can be used to correctly classify a listener as being instructed to attend to the audiobook or to the sounds. In the current paper, we use PS among individuals in the same group to predict the occurrence of interspersed stimuli over time, for each of the three physiological measures. In addition, we investigated if the PS across AA participants was predictive of the occurrence of engaging moments in the book. We aimed to answer the following research questions:

Does PS in EEG and EDA, heart rate and a multimodal metric predict the occurrence of attentionally engaging moments in time? And does this depend on the attentional instruction, type of stimuli and physiological measure?

We expect that interspersed stimulus detection performance of PS measures depends on combinations of the attentional group (AA or SA), the interspersed stimulus type (emotional sounds or beeps) and the physiological measure (EEG and EDA, heart rate or the multimodal metric). We hypothesized the following:

(1) Attentional instruction and stimulus type: (a) for the SA group, detection performance based on PS is above chance for all interspersed stimuli. (b) For the AA group, detection performance based on PS is above chance for emotional sounds, since these attract attention through bottom-up mechanisms related to salience or emotional relevance (Lang, 1995; Öhman et al., 2001; Schupp et al., 2003) irrespective of task instruction. (c) For the AA group, detection performance based on PS is not above chance for beeps, as these are expected to mainly attract attention through top-down mechanisms related to task-instructions.

(2) Physiological measure and stimulus type (a) PS based on peripheral signals (EDA and heart rate) performs better on the detection of emotional sounds than on beeps, because they primarily reflect emotional state (Cacioppo et al., 2007; Boucsein, 2012). (b) PS based on EEG performs better on the detection of beeps than on the detection of emotional sounds, because they primarily reflect top-down selective attention or mental effort (Hogervorst et al., 2014).

(3) Combining physiological measures: combining the physiological measures into a single multimodal metric of PS would result in relatively high detection accuracies when disregarding the differences between stimulus types.

While for the SA group, the timing of short stimuli serve as “ground truth” relevant events to compare to the moments of high PS, we do not know *a priori* what constitutes relevant events or engaging moments in the audiobook. We therefore investigate ratings of post-hoc determined moments of high and low PS in the audiobook by an independent group of participants. We hypothesized that:

(4) Events in audiobook: moments of the audiobook that were associated with high PS in the AA group are rated as more engaging than moments of the audiobook that were associated with low PS.

MATERIALS AND METHODS

Participants

Twenty-seven participants (17 female), between 18 and 48 years old, with an average of 31.6 years and a standard deviation of 9.8 years, were recruited through the institute’s participant pool. Before performing the study, approval was obtained from the TNO Institutional Review Board (IRB). The approval is registered under the reference 2018–70. Prior to the experiment all participants signed informed consent, in accordance with the Declaration of Helsinki. After signing, all participants were randomly assigned to either the AA group or the SA group. After the experiment they received a small monetary reward for their time and traveling costs. None of the participants indicated problems in hearing or attention. Data of one participant were discarded due to failed physiological recordings, resulting in two equal-sized groups.

Materials

EEG, EDA, and electrocardiogram (ECG) were recorded at 1024 Hz using an ActiveTwo Mk II system (BioSemi, Amsterdam, Netherlands). EEG was recorded with 32 active Ag-AgCl electrodes, placed on the scalp according to the 10–20 system, together with a common mode sense active electrode and driven right leg passive electrode for referencing. The electrode impedance threshold was maintained below 20 kOhm. For EDA, two passive gelled Nihon Kohden electrodes were placed on the ventral side of the distal phalanges of the middle and index finger. For ECG, two active gelled Ag-AgCl electrodes were placed at the right clavicle and lowest floating left rib. EDA and heart rate were also recorded using wearable systems (Movisens EdaMove 4 and Wahoo Tickr, respectively). These data are discussed elsewhere (Borovac et al., 2020; Van Beers et al., 2020).

Stimuli and Design

Participants performed the experiment one by one. Each participant was presented with the exact same audio file, composed of a 66 min audiobook (a Dutch thriller “Zure koekjes,” written by Corine Hartman) interspersed with other short, auditory stimuli. Half of the participants were asked to focus on the narrative of the audiobook and ignore all other stimuli or instructions (AA group); and half of the participants were asked to focus on the short, interspersed stimuli and perform accompanying tasks, and ignore the narrative (SA group). The auditory stimuli were 36 emotional sounds, 27 blocks of beeps that SA participants had to keep track of, and an auditory instruction to sing a song. The order of sounds and beeps was randomly determined but was identical for each participant. Inter-stimulus intervals varied between 35 and 55 s, with an average of 45 s and a standard deviation of 6.1 s. We selected these stimuli to evaluate PS for a broad range of events, differing in e.g., audio profile and expected effect on mental processes as a function of task instructions.

Emotional sounds were taken from the second version of the International Affective Digitized Sounds (IADS) (Bradley and Lang, 2007). The IADS is a collection of 6-s acoustic stimuli

that have been normatively rated for valence (positive or negative affect), arousal and dominance. Examples of stimuli are the sound of a crying baby or a cheering sports crowd. We selected 12 neutral sounds (IADS number 246, 262, 373, 376, 382, 627, 698, 700, 708, 720, 723, 728), 12 pleasant sounds (110, 200, 201, 202, 311, 352, 353, 365, 366, 367, 415, 717) and 12 unpleasant sounds (115, 255, 260, 276, 277, 278, 279, 285, 286, 290, 292, 422) based on their normative ratings of valence and arousal. We expected these sounds to attract attention of all participants, even those instructed to ignore the interspersed sounds.

Beeps were presented in blocks of 30 s, with every 2 s a 100 ms high (1 kHz) or low (250 Hz) pitched beep. SA participants needed to separately count the number of high and low beeps presented in a block, as in (De Dieuleveult et al., 2018). This task was practiced with them beforehand. In total, 27 blocks of beeps were presented. We expected these sounds to only attract attention of participants clearly instructed to keep track of them.

Toward the end of the audiobook, the instruction was presented to sing a song aloud after a subsequent auditory countdown reached 0. This instruction had to be followed by the SA group and was expected to induce stress and a strong increase in EDA and heart rate (Brouwer and Hogervorst, 2014). For the current analyses, data following the onset of this stimulus were discarded, because some participants started singing before the counter reached 0. This prohibited analysis of the data in terms of mental processes due to confounding movement effects and artifacts in the data recording.

In total, we consider 3,800 s of data in further analyses, out of which 1,026 s involved concurrent presentation of the audiobook and interspersed stimuli.

Analysis

Pre-processing

Data processing was done using MATLAB 2019a software (Mathworks, Natick, MA, United States). For EEG pre-processing we also used EEGLAB v14.1.2 for MATLAB (Delorme and Makeig, 2004). To remove potentials not reflecting sources of neural activity, but ocular or muscle-related artifacts, logistic infomax independent component analysis (ICA) (Bell and Sejnowski, 1995) was performed. EEG was first down sampled to 256 Hz and high-pass filtered at 1 Hz. This relatively high cut-off frequency has shown to work better for ICA compared to lower cut-off frequencies (Winkler et al., 2015). Data were then notch filtered at 50 Hz, using the standard FIR-filter implemented in EEGLAB function `pop_eegfiltnew`. ICA was performed and the Multiple Artifact Rejection Algorithm (MARA) (Winkler et al., 2011) was used to identify artifactual independent components, i.e., components not reflecting sources of neural activity, but ocular or muscle-related artifacts. These components were removed from re-referenced, but uncleaned data. In these data, samples whose squared amplitude magnitude exceeded the mean-squared amplitude of that channel by more than four standard deviations were marked as missing data (“NaN”) in an iterative way with four repetitions. By doing so, 0.82 % of data were marked as missing.

EDA was downsampled to 32 Hz. The fast changing phasic and slowly varying tonic components of the signal were extracted using Continuous Decomposition Analysis as implemented in the Ledalab toolbox for MATLAB (Benedek and Kaernbach, 2010). In the further analyses we use the phasic component of the signal as this component of the EDA signal is mainly related to responses to external stimuli.

ECG measurements were processed to acquire the inter-beat interval (IBI – inversely proportional to heart rate). After downsampling to 256 Hz, ECG was high-pass filtered at 0.5 Hz. Peaks were detected following Pan and Tompkins (1985). The IBI semi-time series was transformed into a timeseries by interpolating consecutive intervals and resampling at 32 Hz.

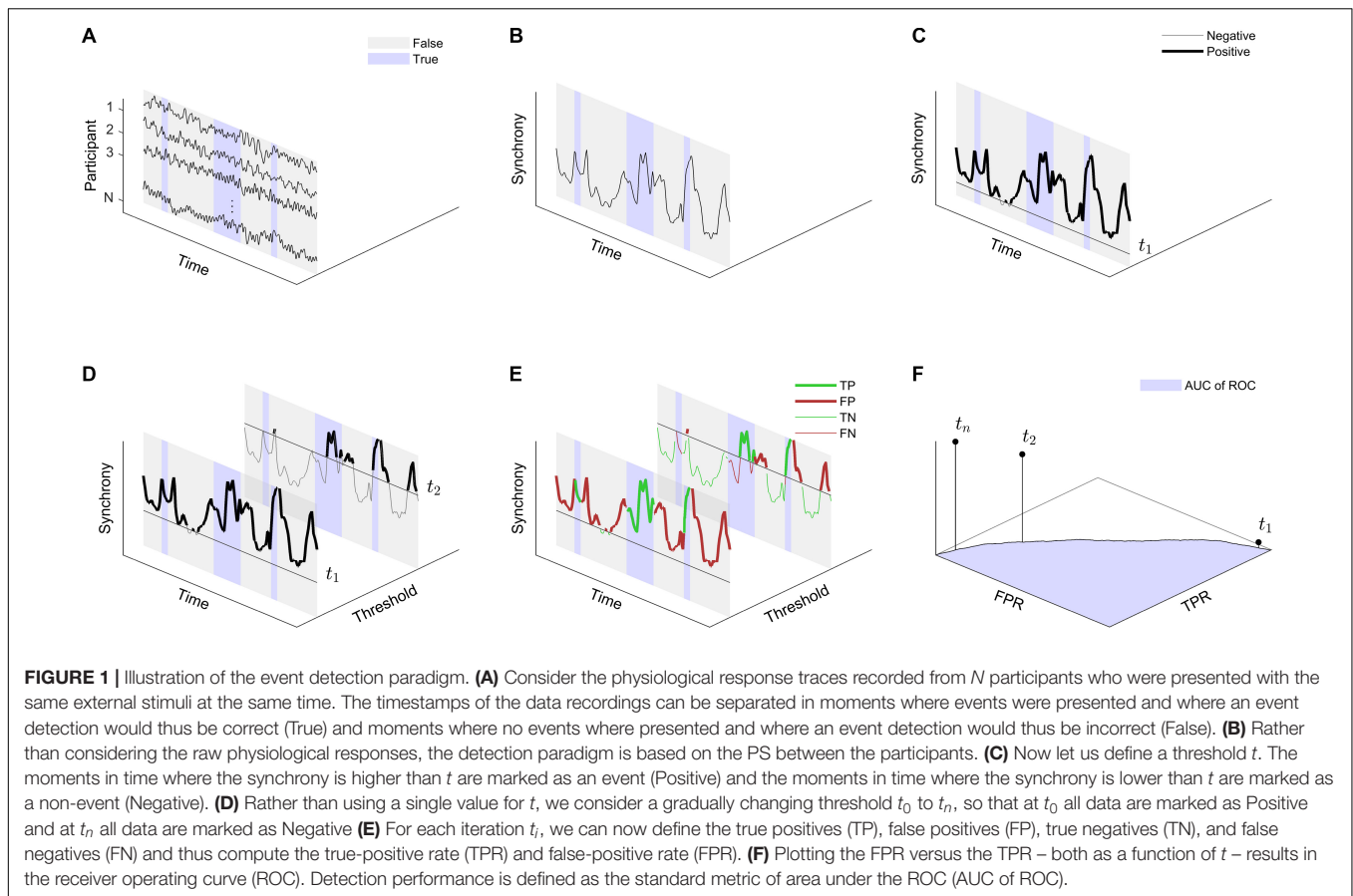
Computation of Inter-Subject Correlations as Measure of Physiological Synchrony

We computed PS by measuring the inter-subject correlations of the neurophysiological signals. For EEG, rather than treating the signals from the 32 channels separately, we evaluated the inter-subject correlations in the correlated components of the EEG (Dmochowski et al., 2012, 2014). The goal of the correlated component analysis is to find underlying neural sources that are maximally correlated between participants, based on linear combinations of electrodes. Components were extracted separately from the AA group and SA group. EEG data from each participant were projected on the component vectors. Participant-to-group inter-subject correlations were then computed as the sum of correlations in the first three component projections, following (Dmochowski et al., 2012, 2014; Cohen and Parra, 2016; Ki et al., 2016; Cohen et al., 2018). Even though we used fewer participants in each attentional group than earlier work on auditory PS (e.g., Cohen and Parra, 2016; Ki et al., 2016), scalp projections of the components were very similar to those obtained in these earlier works, and our EEG PS values were in a similar range of 0.01 to 0.04. For the computation of time-resolved inter-subject correlations, correlations were computed in running 5 s windows at 1 s increments.

Inter-subject correlations in EDA and IBI were computed following (Marci et al., 2007). We computed Pearson correlations over successive, running 15 s windows at 1 s increments as measure of time-resolved inter-subject correlations. Participant-to-group correlations were computed by averaging over all correlations with all other participants in a group.

Physiological Synchrony for the Detection of Interspersed Stimuli

We designed a paradigm to detect relevant events using gradually increasing thresholds to capture the gradual nature of attentional engagement. **Figure 1** provides a visual explanation of our detection paradigm. Consider the EEG, EDA and IBI response traces that were recorded during the experiment. The timestamps of the data recordings can be separated in moments where interspersed stimuli were presented and where an event detection would thus be considered correct (True) and moments where no interspersed stimuli were presented and where an event detection would be considered incorrect (False). Rather than using the raw physiological responses, the detection paradigm is based on the



PS between the participants as a function of time. Now let us define a threshold t . The moments in time where the synchrony is higher than t are marked as an event (Positive) and the moments in time where the synchrony is lower than t are marked as a non-event (Negative). Rather than using a single value for t , we consider a gradually changing threshold t , ranging from the minimum inter-subject correlation value to the maximum inter-subject correlation value. For each iteration of t , we can now define the true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Using this, the true-positive rate or sensitivity (TPR) is then computed as,

$$TPR = \frac{TP}{TP + FN}$$

and the false-positive rate (FPR) or specificity as,

$$FPR = \frac{FP}{FP + TN}$$

Plotting TPR against FPR provides the receiver operating curve (ROC). Detection performance was assessed using the standard metric of the area under the ROC (AUC of ROC).

Chance level performance was assessed using permutations with randomized stimulus timing. In each permutation, the timing of all interspersed stimuli was randomized between the start and the end of the experiment. The same procedure as

above was then applied to obtain the AUC of ROC metric of performance with random stimuli. This procedure was performed on 1000 renditions of such randomized data.

The above-mentioned procedure was repeated $2 \times 3 \times 4$ times, namely for:

- (1) Two attentional groups; considering PS between AA participants and PS between SA participants.
- (2) Three stimulus types; considering as events (True) either blocks of beeps, emotional sounds, or both of these.
- (3) Four physiological measures in which PS is computed; EEG, EDA, heart rate and a multimodal metric that is composed of PS in EEG, EDA and heart rate. To compose this multimodal metric, the PS in EEG, EDA and heart rate were each z-scored. The multimodal PS value at each timestamp was then computed as the average of the z-scored PS values in EEG, EDA and heart rate at that timestamp, for all timestamps ranging from zero to the end of the experiment.

In each condition, one-tailed two-sample t-tests were conducted to test whether detection performance was higher than chance level performance.

Correspondence Between Physiological Synchrony and Reported Engagement With the Audiobook

While for the SA group, the timing of short stimuli served as “ground truth” relevant events to compare to the moments of

high PS, we did not know a priori what constituted relevant, engaging moments in the audiobook. To systematically examine whether moments of high PS were associated with moments of high relevance in the audiobook, we performed a follow-up test in which a second cohort of participants judged clips of the audiobook that were found to be associated with either high or low PS. We recruited 29 participants through the Prolific online experiment environment. All participants signed informed consent before participating. The participants received a small monetary reward for the invested time. We only included participants who indicated to be fluent in Dutch.

We selected clips based on continuous signals of PS among AA participants. We detected the positive and negative peaks in the signals using the 'findpeaks' function in MATLAB. For each measure (EEG, EDA, heart rate and the multimodal metric), the six peaks with highest positive peak-amplitude and six peaks with largest negative peak-amplitude were selected. For each detected peak, we created a 10 s sound clip, that was composed of the 10 s of audio before the detected peak. For four measures, this thus resulted in a total of 48 clips. Clips associated with peaks that were within 10 s of each other were considered to be overlapping, and were merged into one clip by using only the latest of the two clips in time. This resulted in a total of 38 clips that were presented to the participants.

The procedure of the online test was similar to the initial experiment. The participants were first presented with the same audiostream that was presented to the initial cohort of participants. The participants were instructed as participants from the AA group, i.e., to focus their attention on the narrative of the audiobook and ignore any interspersed stimuli as much as possible. After listening to the book, the participants were asked the same questions about the content of the narrative as participants in the initial cohort. We then presented the participants with the sound clips, each of them directly followed by a rating scale. Participants were instructed to rate the preceding clip using an 11-point Likert scale, ranging from 0 to 10. The lower the score, the more the participant's experience corresponded to the words on the left side of the scale (Dutch: 'verveeld', 'kalm', 'ontspannen'; Translated to English: 'bored', 'calm', 'relaxed'). The higher the score, the more the participant's experience corresponded to the words on the right side of the scale (Dutch: 'geïnteresseerd', 'geboeid', 'emotioneel', 'intens'; Translated to English: 'interested', 'fascinated', 'emotional', 'intense'). Using these words, we intended to capture mental states that are expected to be associated with perceiving relevant events, such as engagement, attention and arousal.

For each modality, we tested whether audio clips corresponding to a positive peak in PS were rated as more 'engaging' than audio clips corresponding to a negative peak in PS, using a Wilcoxon signed-rank test. Participants who answered less than three out of ten questions correctly on the questionnaire about the content of the audiobook were considered as not having participated seriously (AA participants in the main experiment answered 5.8 ± 2.0 questions correctly). This concerned three participants. Removing their data left us with data of 26 participants.

RESULTS

Detection of Interspersed Stimuli Using Physiological Synchrony

Figure 2 and Table 1 show our measure of interspersed stimuli detection performance, the AUC of ROC as described in the methods. It is presented separately for AA and SA participants; in EEG, EDA, heart rate, and the multimodal metric; and for blocks of beeps, emotional sounds or both of these stimuli together as to-be identified events. Figure 2 and Table 1 also show the mean and standard deviation AUC of ROC of permutations with randomized event timing as a chance level baseline. Detection performance was largely in line with hypotheses 1 - 3. For the AA group, we found that, as expected, only the occurrence of emotional sounds could be predicted, using PS in EDA ($p < 0.001$) or the multimodal metric ($p = 0.003$). For the SA group, occurrences of beep blocks could be detected well above chance level by PS in EEG, EDA and the multimodal metric ($p < 0.001$, $p = 0.002$ and $p < 0.001$, respectively). The occurrence of emotional sounds could be detected significantly better than chance using PS in EDA, heart rate and the multimodal combination ($p = 0.043$, $p = 0.023$, and $p = 0.011$, respectively). When stimuli were not differentiated according to stimulus type, detection performance was well above chance level for PS in EEG ($p < 0.001$), EDA ($p < 0.001$) and the multimodal metric ($p < 0.001$), but not for PS in heart rate.

Correspondence Between Physiological Synchrony and Reported Engagement

Figure 3 shows engagement ratings of audio clips corresponding to positive peaks and ratings of audio clips corresponding to negative peaks for PS in EEG, EDA, heart rate and the multimodal metric. Results did not follow our hypothesis that audio clips corresponding to positive peaks were rated as more engaging than audio clips corresponding to negative peaks. In fact, in EEG and EDA the opposite effect was found (Wilcoxon test statistic: $W = -3.06$, $p = .002$; $W = -3.44$, $p < .001$, respectively). In heart rate and the multimodal metric no significant difference between ratings corresponding to either positive or negative peaks was found ($W = 0.70$, $p = 0.486$, $W = 0.87$, $p = 0.385$).

DISCUSSION

In the sections below, the hypotheses as stated in the Introduction are discussed separately.

Hypothesis 1: Attentional Instruction and Stimulus Type

We hypothesized that interspersed stimulus detection performance would depend on the attentional group (AA or SA) and the interspersed stimulus type (emotional sounds or beeps), due to bottom-up and top-down mechanisms of attention (Lang, 1995; Öhman et al., 2001; Schupp et al., 2003). For the SA group, we hypothesized that detection performance based on PS would be above chance for all interspersed stimuli, whereas for

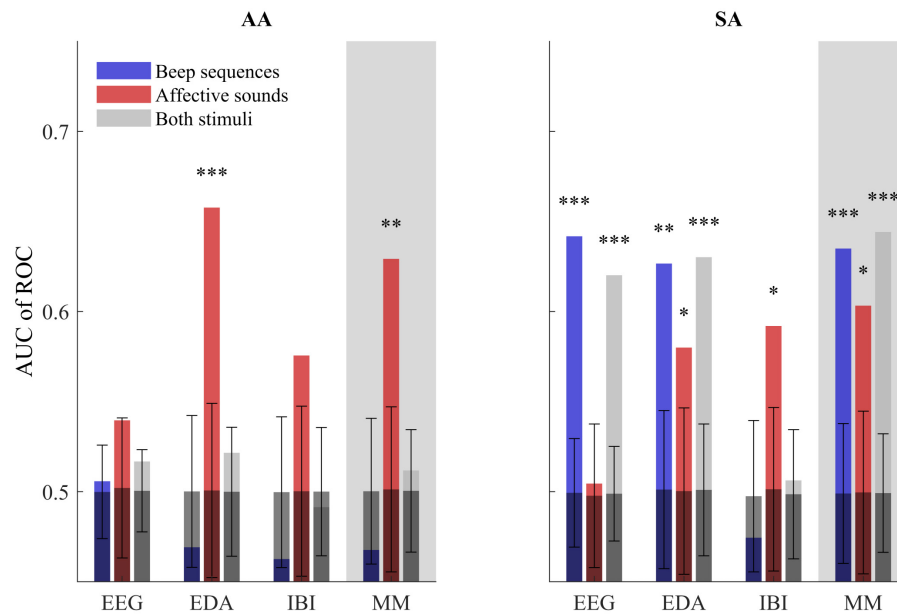


FIGURE 2 | AUC of ROC metric of stimulus detection performance using PS in EEG, EDA, IBI and a multi-modal combination of the three (MM). Performance is shown for the AA and SA groups, when considering only beep blocks or emotional sounds as true positives and when considering both types of stimuli as true positives. In addition, the mean and standard deviation chance level detection performance based on 1,000 renditions with randomized stimulus timing is shown with test results comparing detection performance to chance level (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Note that for the AA group, the emotional sounds but not the beeps are expected to draw (bottom-up) attention, i.e., for the AA group we expect high AUC for emotional sounds only. For the SA group, both beep sequences and emotional sounds are relevant and expected to draw attention.

the AA group we hypothesized that detection performance would be above chance only for emotional sounds, but not for beeps. Results were largely in line with this hypothesis: for the AA group only the emotional sounds were detected with an accuracy above chance level, whereas for the SA group both stimulus types could be detected with above chance level accuracy. Note again that detection performance cannot be directly compared between the different stimulus conditions. There were differences in detection performance between the used physiological measures, these are discussed in the next section.

Hypothesis 2: Physiological Measure and Stimulus Type

Besides the dependency of detection performance on the attentional group and interspersed stimulus type, we expected that detection performance would depend on the used physiological measure and stimulus type. As hypothesized, EEG worked best for the detection of blocks of beeps, while it did not work well for the detection of emotional sounds. We also found EDA to perform well for detecting blocks of beeps. For the detection of emotional sounds, we hypothesized that the peripheral measures (EDA and heart rate) would perform well relative to EEG. Indeed, for both attentional groups, PS in EDA and heart rate perform relatively well for the detection of emotional sounds. Detection performance was significantly above chance using EDA in both groups and using heart rate in the SA group, and near significance ($p = 0.055$) using heart rate in the AA group, whereas detection performance using

PS in EEG was far from significant for emotional sounds in both groups. We think that the observed EEG PS differences between the types of stimuli are the result of both the difference in mental processing (top-down, effortful attention for beeps, versus bottom-up, affective processing for emotional sounds) and low level stimulus features. The beep blocks consisted of precisely-timed, repeated beep occurrences, with constant sound levels, while our emotional stimuli consisted of sounds with irregular sound profiles. The positive results obtained with peripheral measures provide further insight in the mechanisms underlying PS. Whereas previous findings on peripheral PS have been viewed in terms of social relation (Palumbo et al., 2017), we here show that PS in peripheral measures can also be explained by shared attentional engagement. It may be the case that shared attention also underlies results found in contexts of social relation.

Hypothesis 3: Combining Physiological Measures

As the three used physiological measures vary with respect to their ability to reflect different mental states, we hypothesized that combining the physiological measures into a single multimodal metric of PS would result in relatively high detection performance when differences between stimulus types are disregarded. Indeed, the multimodal metric performs best when considering both emotional sounds and beeps as relevant events. Detection accuracies are slightly higher than for the best performing unimodal measure when considering emotional sounds or

TABLE 1 | AUC of ROC metric of stimulus detection performance using PS in EEG, EDA, IBI and a multi-modal combination of the three (MM).

	AA				SA			
	EEG	EDA	IBI	MM	EEG	EDA	IBI	MM
Beep blocks 	0.506 (0.500 ± 0.026) p = 0.410	0.470 (0.500 ± 0.042) p = 0.769	0.463 (0.500 ± 0.042) p = 0.813	0.468 (0.500 ± 0.040) p = 0.790	0.642 (0.499 ± 0.030) p < 0.001	0.627 (0.501 ± 0.044) p = 0.002	0.475 (0.498 ± 0.042) p = 0.708	0.635 (0.499 ± 0.039) p < 0.001
Emotional sounds 	0.540 (0.502 ± 0.039) p = 0.168	0.658 (0.501 ± 0.048) p < 0.001	0.576 (0.500 ± 0.047) p = 0.055	0.629 (0.501 ± 0.046) p = 0.003	0.505 (0.498 ± 0.040) p = 0.433	0.580 (0.500 ± 0.046) p = 0.043	0.592 (0.500 ± 0.045) p = 0.023	0.604 (0.500 ± 0.045) p = 0.011
Both stimuli 	0.517 (0.501 ± 0.023) p = 0.238	0.522 (0.500 ± 0.036) p = 0.273	0.492 (0.500 ± 0.036) p = 0.596	0.512 (0.501 ± 0.034) p = 0.370	0.621 (0.499 ± 0.026) p < 0.001	0.631 (0.501 ± 0.036) p < 0.001	0.507 (0.499 ± 0.036) p = 0.415	0.644 (0.499 ± 0.033) p < 0.001

Performance is shown for the AA and SA groups, when considering only beep blocks or emotional sounds as true positives and when considering both types of stimuli as true positives. In addition, the mean and standard deviation chance level detection performance based on 1,000 renditions with randomized stimulus timing is shown plus p-values of t-tests testing whether detection performance is significantly different from chance level. Gray cells depict conditions in which detection performance is significantly higher than chance level (p < 0.05).

both types of stimuli but not when considering blocks of beeps. We expect that sensor fusion is not beneficial when variables are highly correlated (Hogervorst et al., 2014) – for example for physiological variables all reflecting mental effort – but that sensor fusion can benefit from tasks involving emotional processing besides effortful attentional processing. Besides potentially higher detection performance, the main advantage of multimodal PS seems to be the robustness regarding different types of stimuli, i.e., detection performance varies less between different types of stimuli than for single physiological metrics. In previous work similar effects have been found. For example, when using sensor fusion on machine learning models to distinguish between 13 emotional states, maximum performance was not higher for the multimodal metric, but performance was more robust across the range of emotional states (Verma and Tiwary, 2014). In the end, although adding sensors does not lead to much higher performance compared to the most suitable unimodal recording, a multimodal approach seems to enable detection of relevant events when it is unknown what the best measure for certain stimulus types is. Also note that it is not always known whether certain stimuli will induce mostly effortful cognitive or emotional processing; in many practical cases such processes can co-occur and vary between individuals.

Hypothesis 4: Events in Audiobook

We hypothesized that audio clips corresponding to moments of highest PS would be post-hoc scored as more engaging than audio clips corresponding to moments of lowest PS, but our findings indicated rather the opposite. These findings may be caused by a mismatch between our index of PS and the rating scale of experienced engagement. Post-hoc qualitative analysis of the selected audio-clips revealed that part of the audio clips corresponding to very high PS in EEG coincided with short-term moments of tension or engagement, as expressed through keywords (e.g., swear words) and salient intonation (e.g., a phrase spoken in a very indignant manner). This is in line with earlier work, where moments of high PS in EEG were found to correspond to moments in video clips marked by a high level of short-term suspense, tension or surprise, such as the sight of a gun (Poulsen et al., 2017). Indeed, emotional images and sounds that are rated as highly arousing induce responses in peripheral and central physiological measures (Bradley and Lang, 2000; Lang and Bradley, 2007), which in term may lead to strong PS. In our used audiobook, the keywords that may have driven the particularly high PS contained relatively little important information about the narrative of the story. It seems that this could have been the aspect rated by participants using our engagement scale, leading to a mismatch between self-reported engagement and PS. However, this speculation would need to be investigated further, preferably without having to rely on varying engagement judgements after the fact, but for instance with systematic sentiment analysis (Wöllmer et al., 2013). It is important to further specify what types of attentional engagement can and cannot be captured by PS and how that is dependent on the psychological measure used. Attentional engagement to well-timed events will be better reflected in

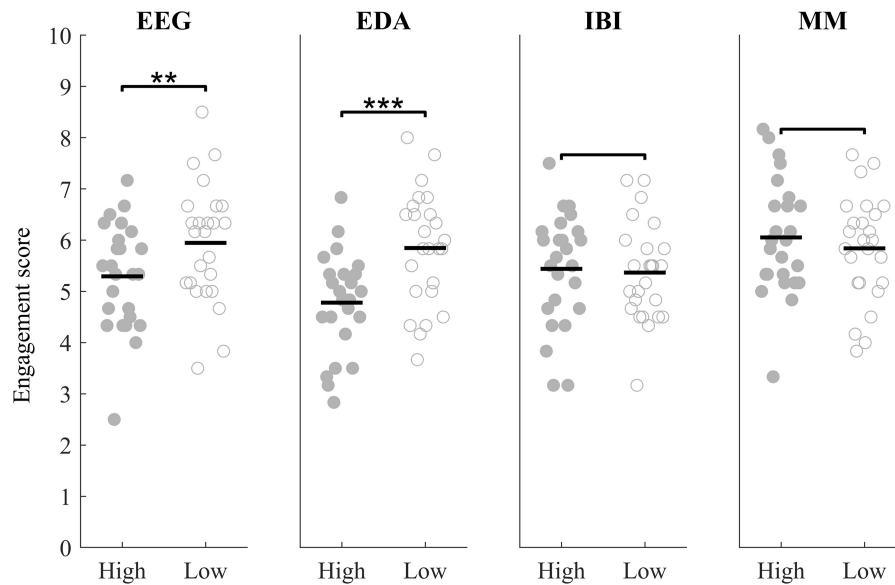


FIGURE 3 | Self-reported engagement scores for audio clips corresponding to moments in the audiobook with high PS (closed markers) or low PS (open markers) in EEG, EDA, IBI and the multimodal metric (MM) (** $p < 0.01$, *** $p < 0.001$).

PS than attentional engagement to less well-timed event on a more abstract level.

Limitations

It should be noted that the stimulus detection performance when not taking stimulus type into consideration ('both stimuli') were mainly driven by detection performance of beep blocks. These beep blocks were interspersed for a total of 810 s, whereas the emotional sounds were only interspersed for a total of 216 s. The stimuli also differed on other aspects. For instance, the beep blocks consisted of precisely timed beeps with immediate stimulus onset equal across trials, whereas the emotional sounds all differed in sound profile. For the AUC of ROC metrics when considering detection of both types of stimuli, beep blocks thus influence the performance metric more than the emotional sounds. While this can be seen as a limitation, this is exemplary for real life situations, where one is interested in detecting relevant, attentionally engaging events, without further specifying or knowing the different types of stimuli, and the proportion of in which they occur; i.e., in such a situation, the 'both stimuli' situation is the default.

We must also note that our simple multimodal approach is certainly not the optimal approach to combine data. In particular, we expect that detection performance can be enhanced by compensating for differences in response latencies across measures. To illustrate the difference in response latency, in response to the same set of emotional sounds, response peak latency ranges from a few 100 ms for EEG to multiple seconds in EDA heart rate (Bradley and Lang, 2000; Hettich et al., 2016). In this paper we simply averaged over response traces in a point-wise fashion, meaning that response-induced peaks may be spread out and their amplitude reduced. Our current results should

therefore be interpreted as a first confirmation that multimodal sensor fusion can be of added value, but we expect that other approaches can greatly enhance performance. In future work we would like to explore other methods for the combination of physiological measures into a multimodal metric of PS.

CONCLUSION

We determined PS in EEG and EDA, heart rate and a multimodal fusion of these three sensors in two groups of participants, that were instructed to attend either to the narrative of an audiobook or to interspersed auditory events. We found that PS could detect the relevant interspersed stimuli with accuracies well above chance level, but also found that moments in the audiobook corresponding to high PS were not rated as more engaging than moments corresponding to low PS. Our results support the notion that PS can be valuable when interested in the course of attentional engagement over time. Currently the relation between PS and engagement is only established for well-defined, interspersed emotional or effortful cognitive stimuli, whereas the relation between PS and a more abstract self-reported metric of engagement is not yet established. We further note that obtained results vary between the used physiological measures. Interesting from a user perspective, EDA worked best overall. These results should enable researchers to monitor PS in situations where intrusive EEG measurements are not suited. However, we also note that the optimal physiological metric may be dependent on the goal of a study and suggest to choose a measure matching with the stimulus of interest. EEG works especially well for well-timed effortful cognitive stimuli, heart rate works especially well for emotional stimuli and EDA works quite well on both types of stimuli. When the stimulus type is unknown, a multimodal

metric may work best as it seems most robust across a broad range of stimuli.

DATA AVAILABILITY STATEMENT

The data presented in this study can be found on the Open Science Framework: <https://osf.io/8kh36/>. The code can be obtained from the corresponding author upon request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the TNO Institutional Review Board. The participants provided their written informed consent to participate in this study.

REFERENCES

- Aliakbarhosseinabadi, S., Kamavuako, E. N., Jiang, N., Farina, D., and Mrachacz-Kersting, N. (2017). Classification of EEG signals to identify variations in attention during motor task execution. *J. Neurosci. Methods* 284, 27–34. doi: 10.1016/j.jneumeth.2017.04.008
- Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129
- Benedek, M., and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* 190, 80–91. doi: 10.1016/j.jneumeth.2010.04.028
- Borovac, A., Stuldreher, I. V., Thammasan, N., and Brouwer, A. M. (2020). “Validation of wearables for electrodermal activity (EdaMove) and heart rate (Wahoo Tickr),” in *Proceedings of 12th International Conference on Measuring Behavior 2020-21*, Kraków, 18–24.
- Boucsein, W. (2012). *Electrodermal Activity*. Cham: Springer Science and Business Media.
- Bradley, M. M., and Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology* 37, 204–215. doi: 10.1111/1469-8986.37.20204
- Bradley, M. M., and Lang, P. J. (2007). *Emotion and Motivation*. Cham: Springer.
- Brouwer, A. M., and Hogervorst, M. A. (2014). A new paradigm to induce mental stress: the sing-a-song stress test (SSST). *Front. Neurosci.* 8:224. doi: 10.3389/fnins.2014.00224
- Brouwer, A. M., Zander, T. O., Van Erp, J. B., Korteling, J. E., and Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Front. Neurosci.* 9:136. doi: 10.3389/fnins.2015.00136
- Cacioppo, J. T., Tassinary, L. G., and Berntson, G. (eds) (2007). *Handbook of Psychophysiology*. Cambridge, MA: Cambridge university press.
- Cohen, S. S., Madsen, J., Touchan, G., Robles, D., Lima, S. F., Henin, S., et al. (2018). Neural engagement with online educational videos predicts learning performance for individual students. *Neurobiol. Learn. Mem.* 155, 60–64. doi: 10.1016/j.nlm.2018.06.011
- Cohen, S. S., and Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *ENeuro* 3:ENEURO.0203-16.2016.
- Davidesco, I., Laurent, E., Valk, H., West, T., Dikker, S., Milne, C., et al. (2019). Brain-to-brain synchrony predicts long-term memory retention more accurately than individual brain measures. *BioRxiv* [Preprint]. doi: 10.1101/644047v1
- De Dieuleveult, A. L., Brouwer, A. M., Siemonsma, P. C., Van Erp, J. B., and Brenner, E. (2018). Aging and sensitivity to illusory target motion with or

AUTHOR CONTRIBUTIONS

A-MB and IS conceived and designed the study. IS and NT performed the experiment. IS analyzed the data with input from A-MB, JE, and NT. IS wrote the first draft of the manuscript. All authors thoroughly reviewed and revised the manuscript drafts and approved the submitted version.

FUNDING

This work was supported by the Netherlands Organization for Scientific Research (NWA Startimpuls 400.17.602).

ACKNOWLEDGMENTS

We thank Ana Borovac for her help with performing the experiment.

- without secondary tasks. *Multisens. Res.* 31, 227–249. doi: 10.1163/22134808-00002596
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., and Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nat. Commun.* 5:4567.
- Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention—a possible marker of engagement? *Front. Hum. Neurosci.* 6:112. doi: 10.3389/fnhum.2012.00112
- Fleureau, J., Guillotel, P., and Huynh-Thu, Q. (2012). Physiological-based affect event detector for entertainment video applications. *IEEE Trans. Affect. Comput.* 3, 379–385. doi: 10.1109/t-affc.2012.2
- Hamadicharef, B., Zhang, H., Guan, C., Wang, C., Phua, K. S., Tee, K. P., et al. (2009). “Learning EEG-based spectral-spatial patterns for attention level measurement,” in *Proceedings of the 2009 IEEE International Symposium on Circuits and Systems*, Taipei, 1465–1468.
- Hanson, S. J., Gagliardi, A. D., and Hanson, C. (2009). Solving the brain synchrony eigenvalue problem: conversation of temporal dynamics (fMRI) over subjects doing the same task. *J. Comput. Neurosci.* 27, 103–114. doi: 10.1007/s10827-008-0129-z
- Hasson, U., Malach, R., and Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* 14, 40–48. doi: 10.1016/j.tics.2009.10.011
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506
- Hettich, D. T., Bolinger, E., Matuz, T., Birbaumer, N., Rosenstiel, W., and Spüler, M. (2016). EEG responses to auditory stimuli for automatic affect recognition. *Front. Neurosci.* 10:244. doi: 10.3389/fnins.2016.00244
- Hogervorst, M. A., Brouwer, A. M., and Van Erp, J. B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Front. Neurosci.* 8:322. doi: 10.3389/fnins.2014.00322
- Hussain, M. S., AlZoubi, O., Calvo, R. A., and D’Mello, S. K. (2011). “Affect detection from multichannel physiology during learning sessions with AutoTutor,” in *Proceedings of the International Conference on Artificial Intelligence in Education*, (Berlin: Springer), 131–138. doi: 10.1007/978-3-642-21869-9_19
- Ki, J. J., Kelly, S. P., and Parra, L. C. (2016). Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *J. Neurosci.* 36, 3092–3101.

- Koole, S. L., Tschacher, W., Butler, E., Dikker, S., and Wilderjans, T. (2020). "In sync with your shrink," in *Applications of Social Psychology*, eds J. P. Forgas, W. D. Crano, and K. Fiedler (Milton Park: Taylor and Francis), 161–184. doi: 10.4324/9780367816407-9
- Lang, P., and Bradley, M. M. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. *Handb. Emot. Elicitation Assess.* 29, 70–73.
- Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *Am. Psychol.* 50:372. doi: 10.1037/0003-066x.50.5.372
- Liu, N. H., Chiang, C. Y., and Chu, H. C. (2013). Recognizing the degree of human attention using EEG signals from mobile sensors. *Sensors* 13, 10273–10286. doi: 10.3390/s130810273
- Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X., and Siipo, A. (2019). Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Comput. Hum. Behav.* 96, 235–245. doi: 10.1016/j.chb.2018.06.030
- Marcí, C. D., Ham, J., Moran, E., and Orr, S. P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *J. Nerv. Ment. Dis.* 195, 103–111. doi: 10.1097/01.nmd.0000253731.71025.fc
- Öhman, A., Flykt, A., and Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *J. Exp. Psychol. Gen.* 130:466. doi: 10.1037/0096-3445.130.3.466
- Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., and Goodwin, M. S. (2017). Interpersonal autonomic physiology: a systematic review of the literature. *Pers. Soc. Psychol. Rev.* 21, 99–141. doi: 10.1177/1088868316628405
- Pan, J., and Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* 32, 230–236. doi: 10.1109/tbme.1985.325532
- Perez, P., Madsen, J., Banellis, L., Turker, B., Raimondo, F., Perlberg, V., and Similowski, T. (2020). Conscious processing of narrative stimuli synchronizes heart rate between individuals. *bioRxiv* doi: 10.1101/2020.05.26.116079
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., and Hansen, L. K. (2017). EEG in the classroom: synchronised neural recordings during video presentation. *Sci. Rep.* 7:43916.
- Rosenberg, M., Noonan, S., DeGutis, J., and Esterman, M. (2013). Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task. *Attent. Percept. Psychophys.* 75, 426–439. doi: 10.3758/s13414-012-0413-x
- Schupp, H. T., Junghöfer, M., Weike, A. I., and Hamm, A. O. (2003). Attention and emotion: an ERP analysis of facilitated emotional stimulus processing. *Neuroreport* 14, 1107–1110. doi: 10.1097/00001756-200306110-00002
- Stuldreher, I. V., Thammasan, N., Van Erp, J. B. F., and Brouwer, A. M. (2020). Physiological synchrony in EEG, electrodermal activity and heart rate reflects shared selective auditory attention. *J. Neural Eng.* 17:046028. doi: 10.1088/1741-2552/aba87d
- Tickle-Degnen, L., and Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychol. Inq.* 1, 285–293. doi: 10.1207/s15327965pli0104_1
- Tourunen, A., Kykyri, V. L., Seikkula, J., Kaartinen, J., Tolvanen, A., and Penttonen, M. (2019). Sympathetic nervous system synchrony: an exploratory study of its relationship with the therapeutic alliance and outcome in couple therapy. *Psychotherapy* 57, 160–173. doi: 10.1037/pst0000198
- Van Beers, J., Stuldreher, I. V., Thammasan, N., and Brouwer, A. M. (2020). "A comparison between laboratory and wearable sensors in the context of physiological synchrony," in *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI)*, Utrecht, 604–608.
- Verma, G. K., and Tiwary, U. S. (2014). Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage* 102, 162–172. doi: 10.1016/j.neuroimage.2013.11.007
- Winkler, I., Debener, S., Müller, K. R., and Tangermann, M. (2015). "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP" in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, 4101–4105.
- Winkler, I., Haufe, S., and Tangermann, M. (2011). Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behav. Brain Funct.* 7:30. doi: 10.1186/1744-9081-7-30
- Wöllmer, M., Weninger, F., Knaup, T., Schuller, B., Sun, C., Sagae, K., et al. (2013). Youtube movie reviews: Sentiment analysis in an audio-visual context. *IEEE Intelligent Syst.* 28, 46–53. doi: 10.1109/mis.2013.34
- Zander, T. O., and Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005
- Zehetleitner, M., Goschy, H., and Müller, H. J. (2012). Top-down control of attention: it's gradual, practice-dependent, and hierarchically organized. *J. Exp. Psychol.* 38:941. doi: 10.1037/a0027629

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Stuldreher, Thammasan, van Erp and Brouwer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.