I.V. Stuldreher

Delft University of Technology

Monitoring selectively directed auditory attention using physiological synchrony in EEG, electrodermal activity and heart rate

I.V. Stuldreher

TU Delft TNO

TNO TU Delft

Monitoring selectively directed auditory attention using physiological synchrony in EEG, electrodermal activity and heart rate

by

I.V. Stuldreher

to obtain the degree of Master of Science in Mechanical Engineering
at the Delft University of Technology,
to be defended publicly on Wednesday July 24, 2019 at 08:30.

Student number:	4353455	
Project duration:	September 1, 2018 - July 31, 2019	
Thesis committee:	Dr. ir. J. C. F. de Winter,	Delft University of Technology
	Dr. D. Dodou,	Delft University of Technology
	ir. Y. B. Eisma,	Delft University of Technology
	Dr. L. Peternel,	Delft University of Technology
	Dr. A-M. Brouwer,	Netherlands Organisation for Applied Scientific Research (TNO)

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

TNO

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

On the cover: Students wearing a portable electroencephalogram (EEG) in a classroom to monitor synchrony in brain activity between students. © Diane Quinn

Abstract

Monitoring selectively directed auditory attention in groups can be helpful in a range of contexts, such as in education. In real-world settings, selectively directed attention cannot be monitored by relating physiological signals to known event markers. Determining the similarity of physiological responses across individuals in a group (i.e., physiological synchrony - PS) may provide a solution to this problem, as it has been shown reflective of shared attentional engagement toward audiovisual stimuli. This study was aimed at examining whether PS in autonomic and neural measures are suitable markers to recover selectively directed attentional focus and moments of emotional or task-related relevance in a composite auditory stimulus.

Electroencephalography (EEG), electrodermal activity (EDA) and cardiac inter-beat interval (IBI) were monitored from participants who heard a composite auditory stimulus, consisting of a narrative audiobook, interspersed with short stimuli. One group of the participants ($n = 13$) was instructed to attend to the narrative, and the other group ($n = 13$) was instructed to attend to the short stimuli, that were tones that attending participants needed to keep track of, emotional sounds and a stress test attending participants needed to perform.

EEG and EDA signals of participants were more strongly synchronized with those of participants in the same attentional condition than with those of participants in the other attentional condition. No such effect was found in IBI. For a single individual, PS in EEG allowed attribution to the correct attentional group in 85% of the cases, for EDA this was 81% and for IBI accuracy was below chance level. Further analyzing EEG, PS was higher during relevant stimulus presentation than over the entire audiobook. PS across stimulus-attending participants was higher than across book-attending participants during these short stimuli.

It is concluded that PS in both autonomic and neural measures can be suitable to recover selectively directed attention, as indicated by the high classification accuracy for EEG and EDA. The results of this study also indicate that PS may be suited for the detection of moments of emotional or task-related relevance, as PS in EEG was higher during relevant stimulus presentation than over the entire experiment. Future research should investigate whether PS in autonomic measures could also be suited for detection of relevant events.

Acknowledgments

This thesis marks the end of my master in Mechanical Engineering, in which I followed the BioMechanical Design program. During this master program, my interests were gradually drawn towards research monitoring and enhancing human experience while interacting with machines or computers. This is no field for engineers alone; we as engineers should collaborate with psychologists, human-factors specialists and many more. That is why I was very eager to start an internship in an interdisciplinary project focused on mental state monitoring using physiological measures at the Department of Perceptual and Cognitive Systems of TNO in September 2018. This internship enabled me to conduct human-subject research for this thesis, to experience in-the-field experiments, to write an accepted conference paper and eventually resulted in the thesis presented here. I also learned some new things about myself. Even though the scientific community always had some attraction to me, I was never sure if I wanted to pursue a career in scientific research. Now, this has changed. I would definitely like to contribute to the scientific community. A good start is to write a paper from the research presented in this thesis, which I would like to submit for publication.

All of this would not have been possible alone. I would hereby like to acknowledge a few people who supported me during this research. First and foremost, I would like to thank my supervisors Anne-Marie Brouwer and Joost de Winter. Anne-Marie, thanks a lot for giving me the opportunity to work on this project, for enabling me to frame a research scope suited for me and for guiding me through this research when necessary. Your unconditional positive attitude always energized me to get the most out of this project. Joost, thank you for your genuine interest in my work and all the honest, critical and thorough feedback during the entire project. You always encouraged me to get the most out of my work, for which I am grateful.

Natty and Ana, thanks a lot for helping me collecting the data and for all the fruitful brainstorming sessions about the analysis of the physiological data. Thanks to all the interns at TNO, for the support, for sharing highs and lows of our internships, but mainly for making the time at TNO fun. Thanks to my friends and family for their unconditional support, encouragement and interest in my progress. Fenna, special thanks to you, for all your support and for offering a listening ear when I needed one.

Contents

1

Introduction

1.1. Selectively directed attention and physiological responses

Monitoring selectively directed attention can be helpful in a range of contexts, for example to study and support children in the classroom who suffer from attentional problems. Continuous and implicit measures of attention may be extracted from physiological signals, such as changes in heart rate (HR) or electrodermal activity (EDA), reflective of autonomic nervous system (ANS) activity, or changes in responses of central nervous system (CNS) activity, such as electroencephalographic brain potentials (EEG). For instance, asking observers to attend to one out of a range successively presented stimuli generates a P300-peak in EEG (Polich and Kok, 1995), emotional sounds heighten EDA (Bradley and Lang, 2007) and a mental working memory related task induces responses in all of these measures (Hogervorst et al., 2014). However, in all of these examples, physiological measures need to be related to the time that stimuli of interest occur. In real-life contexts, this is usually difficult to do from a practical point of view. In addition, it is often not clear what the stimulus of interest is.

1.2. Physiological synchrony as a measure of shared attention

A promising approach that could deal with these problems is to determine the similarity of physiological responses across individuals in a group (i.e., physiological synchrony - PS). Using narrative movie or audio clips, neural responses were shown to be reliably reproduced across individuals with shared attentional focus (Dmochowski et al., 2012; Hanson et al., 2009; Hasson et al., 2010, 2004). Moments of high inter-subject correlations (ISC) across electroencephalographic responses correlated with arousing moments of stimulus presentation and predicted audience preferences during popular television content (Dmochowski et al., 2014, 2012). ISC were also found to be higher when participants naturally attended to narrative stimuli than when they were performing mental arithmetic tasks during stimulus presentation (Ki et al., 2016). Promising for real-world studies, many of these results were reproduced with portable and low-cost equipment (Poulsen et al., 2017). Note that joint physiological change across individuals as measure of shared attention to naturalistic stimuli has mainly been investigated for neural responses. Only a few exceptions used PS in autonomic responses, such as Marci (2006), who asso-

ciated autonomic PS with shared attentional engagement toward TV commercials. Autonomic PS has however intensively been studied in contexts of affective connectedness and social processes, recently reviewed by Palumbo et al. (2017). PS in ANS activity has for example been associated with relationship quality of romantic couples, empathy in parent-child dyads and team-performance of team mates (Elkins et al., 2009; Levenson and Gottman, 1983; Marci et al., 2007). These findings may also be explained through mechanisms of shared attention and emotional experience, as heightened autonomic arousal also associates with enhanced attention (Critchley, 2002). In earlier work (Stuldreher et al., submitted, added as supplementary material in Appendix B), we found that PS in EEG and autonomic measures have not yet concurrently been monitored, so that it is still unclear to what extent PS in ANS activity can be reflective of shared attention compared to PS in neural activity.

1.3. Aims

Up to now, PS showed to be a reliably marker of shared attention, but the effect of directing attention to specific stimulus aspects on PS across participants is still unknown. The current study was aimed at examining whether PS in autonomic and neural measures are suitable markers to identify selectively directed attention and moments of emotional or task-related relevance of individuals who are all presented with the same stimulus, and are all attending to it, be it to different stimulus aspects. EEG, EDA and IBI were monitored from participants that were presented with the same, composite auditory stimulus, consisting of a narrative audiobook, interspersed with short stimuli. Half of the participants ($n = 13$) were instructed to attend to the narrative, and the others ($n = 13$) were instructed to attend to the short stimuli, that were tones that attending participants needed to keep track of, emotional sounds and a stress test attending participants needed to perform.

To obtain an indication of the suitability of each of the physiological responses in distinguishing between groups with different attentional instructions, response traces of EEG, EDA and IBI time-locked to the short stimuli were obtained for both groups of participants. It was hypothesized that response deflections would be larger for short-stimuli-attending (SSA) participants, than narrative-attending (NA) participants, but that this effect would differ between types of stimuli, as some stimuli involuntarily draw attention of all participants (emotional sounds), while other stimuli are very dependent on task instructions (sequences of tones and the stress test). It was then investigated whether EEG, EDA and IBI signals of participants more strongly synchronized with those of participants in the same attentional condition than with those of participants in the other condition. Aiming to identify directed attention, the attentional condition of each participant was classified as indicated by the group with whom he or she showed the highest average PS for each of the physiological measures. EEG was then further analyzed, by investigating the effect of interspersed stimulus presentation on PS in EEG of both attentional groups, as a first step toward identification of emotionally or task-related relevant events in real-world environments.

2

Methods

2.1. Participants

Twenty-seven participants (17 female, 10 male) took part in the experiment. Age ranged from 18-48, (M = 31.6, SD = 9.8 years). All participants indicated to have normal hearing and indicated to have no known attention problems, such as attention deficit (hyperactivity) disorder. The study was approved by the TNO Institutional Review Board (TCPE) as well as the TU Delft Human Research Ethics Committee. All participants signed informed consent prior to the experiment, in accordance with the Helsinki Declaration of 1975 as revised in 2014 (World Medical Association, 2014). Participants were randomly assigned in one of two experimental groups. Data from one participant was discarded due to failed physiological recordings, resulting in 13 participants per group. Participants received a small monetary award to make up for their time and travel.

2.2. Stimuli and design

Participants were all presented with a naturalistic narrative stimulus - a 66-minute audio book (Zure Koekjes, written by Corine Hartman, read by Willemijn de Vries) - interspersed with short cognitive and affective auditory stimuli. Independent variables were attentional condition and short-stimulus type. Half of the participants were asked to focus on the narrative of the audiobook and ignore all other stimuli or instructions (narrative-attending - NA) and the other half of the participants were asked to focus on the short stimuli and perform accompanying tasks, and ignore the narrative (short-stimuli-attending - SSA). The set of stimuli short stimuli consisted of 36 sounds from the second version of the International Affective Digital Sounds (IADS-2) collection (Bradley and Lang, 2007), 26 cognitive working-memory tasks (WMT) adopted from De Dieuleveult et al. (2018) and a shortened, auditory version of the sing-a-song stress test (SSST) (Brouwer and Hogervorst, 2014).

The IADS-2 is a set of acoustic stimuli with normative three dimensional ratings of emotion (i.e., valence, arousal and dominance), provided on a 9-point scale. Sounds were selected based on their ratings of valence and arousal. Three sets of 12 sounds each were selected: neutral (valence: $M = 5.9$, $SD = 1.7$; arousal: $M = 5.4$, $SD = 1.6$), pleasant (valence: $M = 7.0$, $SD = 0.4$; arousal: $M = 6.8$,

$SD = 0.4$) and unpleasant (valence: $M = 1.8$, $SD = 0.2$; arousal: $M = 7.5$, $SD = 0.4$). See Section C.1 for the normative ratings of valence, arousal and dominance of the individual sounds. The duration of all sounds was 6 seconds.

The WMT consisted of a sequence of low (250 Hz, 100 ms) and high (1 kHz, 100 ms) tones, of which one was presented every two seconds for a total duration of 30 seconds. For each sequence of tones, short-stimuli attending participants needed to separately count the number of low and high tones.

The SSST is a stress test originally developed for visual presentation, where participants read the instruction to sing a song aloud after having read multiple neutral sentences. In the current experiment, participants were presented with an auditory version of the test. 106 seconds before the audiobook ended, participants heard (translated from Dutch): 'I will count down from ten to zero. At zero, start singing a song out loud. Ten. Nine. Eight. ... One. Zero.' The duration of the total instruction was 26.7 seconds. The duration of the countdown interval was 18.7 seconds. If participants started singing, they were told to stop singing.

All participants were presented with the exact same audio stream, consisting of the naturalistic stimulus and the interspersed stimuli, with the same inter-stimulus intervals and order. Inter-stimulus intervals varied between 35 and 55 seconds ($M = 45$, $SD = 6.1$), to ensure unpredictable stimulus onset. The exact stimulus order and intervals can be found in Section C.3. The audio streams were combined offline and played as one stream of audio online. The audio stream was presented using standard computer speakers, set at maximum volume. Windows audio volume was set at 20 %.

2.3. Physiological measurements

Electrocardiogram (ECG), EDA and EEG were all recorded using the ActiveTwo system (BioSemi, Amsterdam, Netherlands). ECG was recorded with two active Ag-AgCl electrodes placed at the right clavicle and lowest floating left rib. EDA was recorded using two passive Nihon Kohden electrodes placed on the ventral side of the distal phalanges of the middle and index finger at the non-dominant hand. 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 (CMS) active electrode and a driven right leg (DRL) passive electrode for referencing. Electrode impedance threshold was set at 20 kOhm, electrodes with impedances exceeding this threshold were re-attached. Recording frequency was 1024 Hz.

EDA and HR were also recorded using wearables for the purpose of validation - these data were analyzed in another study (Borovac et al., submitted). The used HR wearable was a Wahoo Tickr (Wahoo Fitness, Atlanta, GA, USA), for which a band was fitted around the chest after applying gel on its sensors. The used EDA wearable was an EdaMove 4 (Movisens GmbH, Karlsruhe, Germany), for which two self-adhesive electrodes were placed on the palm of the non-dominant hand.

2.4. Procedure

Participants were seated in the experimental room. The experimental room did not have outside windows and had constant office lighting. After reading information about the study and signing the informed consent, both of which can be found in Appendix A, the physiological sensors were attached. First the HR wearable and EDA wearable, followed by the ECG, EDA and EEG sensors. Figure 2.1 shows a participant seated in the room in which the experiment was conducted, equipped with physio-

Figure 2.1 A participant in the experimental setup. EDA sensors were placed on the non-dominant left hand, ECG sensors were placed on the lower left rib and right clavicle and a 32 channel EEG-cap was placed according to the 10-20 system. Audio was played through the white speakers placed in front of the participant.

logical sensors.

Participants were instructed that a 70-minute audio stream would be presented, consisting of an audiobook, interspersed with short auditory stimuli. All participants were told that three stimulus categories were presented, namely emotional sounds, sequences of low and high tones and a spoken assignment toward the end of the audiobook. Participants were then instructed to which information they should selectively attend depending on their experimental group (see Section 2.2 and Appendix A).

All participants were then instructed to sit still, not to exert pressure on the EDA electrodes, to keep their eyes open and to be quiet during the experiment, unless otherwise instructed. Physiological recordings were then started. Baseline recordings were conducted for 60 seconds without auditory stimuli, and 60 seconds of white noise. The experiment was then started. Participants who started singing during the sing-a-song stress test were instructed that they could stop singing. After the experiment, baseline recordings were again conducted during 60 seconds of white noise, followed by 60 seconds without auditory stimulation.

Afterwards, participants were presented with two questionnaires (see Appendix D). The first questionnaire was used to obtain self-reported measures of attention, workload and emotional experience. The second questionnaire contained questions about the narrative of the audiobook and the short stimuli and was used to obtain a performance metric indicative of directed attention towards the audiobook and the short stimuli.

2.5. Analysis

2.5.1. Preprocessing

Data processing was done offline using MATLAB 2018b software (Mathworks, Natick, MA, USA). A link to the Matlab scripts used for this study can be found in Section E.3. ECG measurements were

processed to acquire the inter-beat interval (IBI). After downsampling to 256 Hz to limit necessary computational resources, ECG was high-pass filtered at 0.5 Hz. Peaks were detected from a squared version of the reconstructed frequency-localized version of the ECG waveform using wavelets (The Mathworks, Inc., 2015). In Section E.1 this process is explained in more detail. The R-to-R interval (i.e., IBI) was extracted. The IBI semi-time series was transformed into a time-series in order to compare data between participants. This was done by interpolating consecutive IBIs and then resampling at 2 Hz. An impression of the data obtained for each participant can be found in Section F.1, in which IBI and raw electrodermal responses for each individuals are shown.

EDA was first downsampled to 64 Hz. For analysis of time-locked stimulus response traces, the phasic (SCR) and tonic (SCL) component of the electrodermal response were extracted using Continuous Decomposition Analysis (Benedek and Kaernbach, 2010) as implemented in the Ledalab toolbox for MATLAB. For analysis of stimulus-locked responses, data were standardized into z-scores (i.e., mean of 0 and standard deviation of 1), following Ben-Shakhar (1985), as:

$$z(t) = \frac{x(t) - \bar{x}}{\sigma} \tag{2.1}$$

where $x(t)$ is the phasic component at time t, with mean $\bar{x}$ and standard deviation σ of all phasic data points obtained during the experiment, including pre- and post-experiment baselines.

EEG was processed offline with EEGLAB v14.1.2 for MATLAB (Delorme and Makeig, 2004). EEG was first downsampled to 256 Hz - again to limit necessary computational resources - high-pass filtered at 1 Hz and notch filtered at 50 Hz, using the standard FIR-filter implement in EEGLAB function `pop_eegfiltnew`. Channels were re-referenced to the average channel values. Note that re-referencing is a must for data obtained with the BioSemi ActiveTwo system, as the raw signals still contain some common-mode noise. Logistic infomax independent component analysis (ICA, Bell and Sejnowski, 1995) was performed on more strongly filtered data (see Section E.2) to classify artifactual independent components, i.e., components not reflecting sources of neural activity, but ocular or muscle-related artifacts. These components were removed from the 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').

2.5.2. Self-report measures

After stimulus presentation, self-report measures of experience and measures indicative of attentional performance were obtained. Self-reported measures of attention, workload and emotional experience (see Section D.1) were provided on a continuous scale from 'not at all' to 'very', which were mapped to a value from 0 to 1. Between-group differences for these self-reported measures were tested with independent-sample t-tests.

As a measure of directed attention, participants were asked to answer questions about the short, interspersed stimuli as well as the content of the audiobook (see Section D.2). For each participant, the number of correctly answered questions about the audiobook content was computed. Between-group differences were tested for significance with a Wilcoxon Rank Sum test. Similar, between group differences were tested for significance for the number of correctly described affective sounds, and for

the difference between the reported average number of tones and the correct answer. One participant was excluded from the analysis of affective sound answers due to familiarity with the presented affective sounds.

2.5.3. Stimulus-locked physiological response traces

To obtain an initial impression of the suitability of each of the physiological measures in distinguishing between groups with different attentional instructions, response traces time-locked to the onset of each of the types of short stimuli were extracted. Neural event-related potentials were obtained from frontal (Fz), central (Cz) and parietal sites (Pz) on the anterior-posterior midline of the scalp. Pre-processed neural potentials were cut in 2000 ms stimulus-locked epochs (100 ms pre-stimulus onset, 1900 ms post-stimulus onset) and baseline corrected from -100 to 0 ms relative to stimulus onset. For each participant, response traces were averaged over all trials of each stimulus type. Neural event-related potentials were not obtained in response to the SSST, as it has no clear time-locked stimulus onset and only one trial was presented to each participant. For assessment of neural activity during the WMT, responses were first averaged over all tones presented in each sequence, before averaging over trials. Grand-average stimulus-locked potentials were obtained by averaging over all participants within each group. The standard-error of the mean (SEM) across participants in each group at time t was computed as the standard deviation across participants at time t divided by the number of participants. Running independent-sample t-tests were conducted to test for significant between-group differences over time. Tests were adjusted for multiple comparisons by controlling the false discovery rate (FDR) using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). In this procedure, p-values are sorted and ranked. The smalles value gets rank 1, the largest rank N. All p-values are then multiplied by N and divided by their rank to obtain adjusted p-values. The FDR threshold was set to 0.05, so that adjusted p-values smaller than 0.05 were selected as significant.

Autonomic response traces (phasic EDA and IBI) were cut in 31 s epochs (1 s pre-stimulus onset, 30 s post-stimulus onset) and baseline corrected from -1 to 0 s relative to stimulus onset. For each stimulus type, time-locked responses were averaged over trials within each participant and averaged over all participants in each group to obtain grand-average group responses. Autonomic response trace between-group differences were tested for significance with running t-tests and adjusted for multiple comparisons using the procedure presented above.

2.5.4. Physiological synchrony

Correlated component analysis for assessment of neural synchrony

Similarity of neural responses between participants in the time-domain was assessed using correlated component analysis (CorrCA) (Dmochowski et al., 2012). CorrCA is similar to the more familiar principal component analysis, except that projections of CorrCA capture maximal correlations between data sets instead of maximal variance within a set of data. CorrCA is conceptually similar to canonical correlation analysis (Hotelling, 1936), differing only in that it uses the same projections for all data sets. Figure 2.2, obtained from Ki et al. (2016), illustrates how CorrCA is used to assess neural reliability.

For a given stimulus presented to N participants, there is a set of N data sets $\{X_1, \dots X_N\}$ with $X_i \in \mathbb{R}^{D \times T}$, where D is the number of channels and T is the number of samples in time. CorrCA requires the

calculation of the pooled between-subject cross-covariance:

$$R_b = \frac{1}{N(N-1)} \sum_{k=1}^{N} \sum_{l=1, l \neq k}^{N} R_{kl} \tag{2.2}$$

and pooled within-subject covariance:

$$R_w = \frac{1}{N} \sum_{k=1}^{N} R_{kk} \tag{2.3}$$

where

$$R_{kl} = \sum_{t=0}^{T} (X_k(t) - \bar{X}_k)(X_l(t) - \bar{X}_l)^T \tag{2.4}$$

measures the cross-covariance of all channels from participant k with all channels from participant l. $X_k(t)$ represents the D channel values at time t in subject k and $\bar{X}_k$ represents their mean value in time. We seek to find a projection vector v_i that captures the strongest correlation between datasets. These components are the eigenvectors of matrix $(R_w^{-1} R_b)$, with strongest eigenvalues λ_i, as follows:

$$(R_w^{-1} R_b) v_i = \lambda_i v_i \tag{2.5}$$

Before computing the eigenvectors, the pooled within-subject correlation matrix was regularized to improve robustness to outliers using shrinkage (Blankertz et al., 2011), as

$$R_w \leftarrow (1-\gamma) R_w + \gamma \bar{\lambda} I \tag{2.6}$$

where $\bar{\lambda}$ corresponds to the mean eigenvalue of R_w and γ is the shrinkage parameter, here selected to be 0.5. To determine reliability of EEG responses of a single participant with respect to the group, data are projected onto the component vectors, as

$$y_{ik}(t) = v_i^T (x_{ik}(t) - x_k) \tag{2.7}$$

Participant-to-group correlations is then computed as the sum of ISC with all other participants summed over the first three components, as

$$\bar{r}_k = \sum_{i=1}^{3} r_{ki} \tag{2.8}$$

where

$$r_{ki} = \frac{1}{N(N-1)} \sum_{l=1, l \neq k}^{N} \frac{\sum_{t=0}^{T} y_{ik}(t) y_{il}(t)}{\sqrt{\sum_{t=0}^{T} y_{ik}(t)^2 \sum_{t=0}^{T} y_{il}(t)^2}} \tag{2.9}$$

are the Pearson correlation coefficients, averaged across pairs of participants and computed for compo-

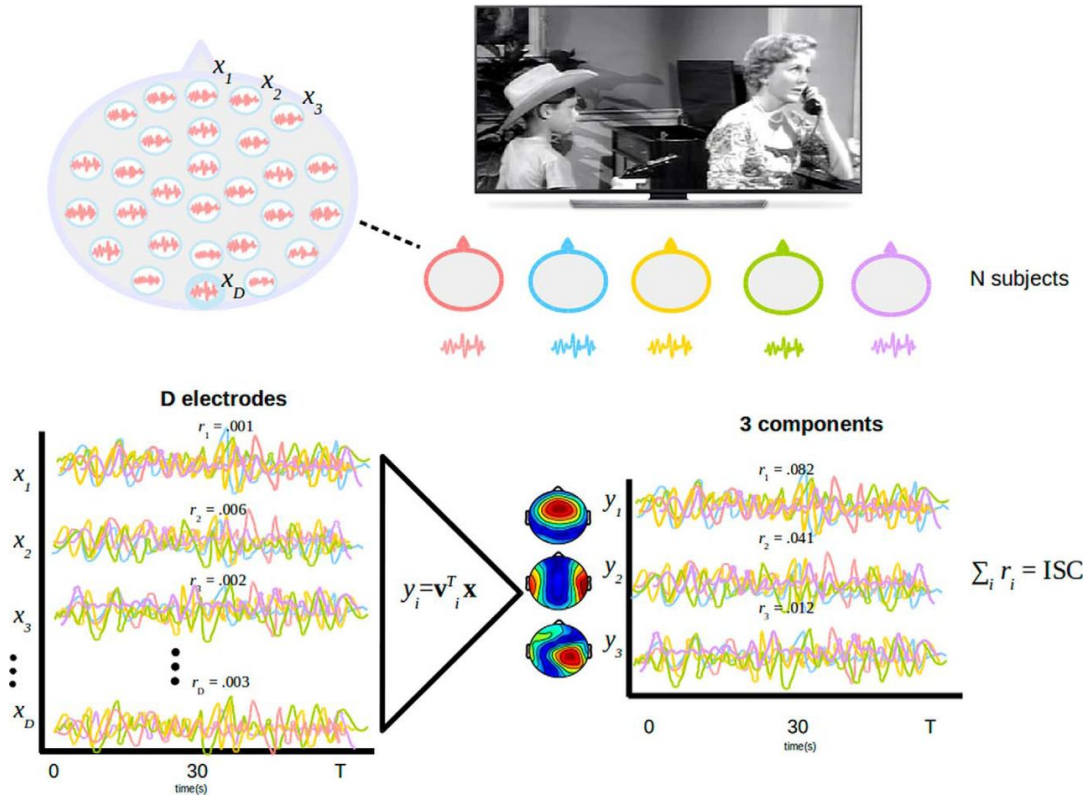


Figure 2.2 Overview of CorrCA. Physiological response are recorded from N participants during presentation of a naturalistic stimulus. Each subject provides time-series $X(t)$ recorded on D channels. The data are projected using projection vectors, v_i, that maximize correlations, r_i. ISC is measured as the sum of correlation of the first few - in this case three - correlated components. Obtained from Ki et al. (2016).

nent projections. The scalp projections of the correlated components (A) were obtained following Parra et al. (2005),

$$A = R_w v_i (v_i^T R_w v_i)^{-1} \tag{2.10}$$

Chance level ISC were determined using phase randomized surrogate data. Phase of the EEG time-series was randomized in the frequency domain, following Theiler et al. (1992). By using this method, new time-series of EEG responses are obtained, in which the temporal fluctuations are no longer aligned in time with the original data and thus do not correlate across subjects. Chance-level correlations still remain, as these depend on the spatiotemporal correlation in the data. After randomization, the same procedure as described above was followed. Obtained ISC could then be compared to the mean correlations from 10 phase-randomized data sets.

Inter-subject autonomic synchrony

Physiological synchrony in autonomic responses was quantified using a moving window approach, introduced by Marci et al. (2007) for the assessment of PS in electrodermal responses over time. In the current study, the phasic component of the electrodermal response and IBI were used for assessment of PS. Pearson correlations were calculated over successive, running 15 s windows at 1 s increments. The overall correlation between two responses was computed as the natural logarithm of the sum of all

positive correlations over the sum of the absolute values of all negative correlations, as

$$r_{kl} = \log\left(\frac{\sum_{t=0}^{T} r_{kl}(t) > 0}{\sum_{t=0}^{T} |r_{kl}(t) < 0|}\right) \tag{2.11}$$

where $r_{kl}(t)$ are the correlations between individual k and l in the moving window at time t.

2.5.5. Subject-to-group correlations for identifying attentional focus

Neural subject-to-group correlations

To discriminate between task conditions, correlated component vectors v_i were extracted separately from the two groups of participants. Data of each participant were projected on these component vectors and correlations between a to-be-classified participant with all other members of the group where component vectors were extracted from, were computed. The average between the correlations of a to-be-classified participant with all participants in the group where components were extracted from is the resulting subject-to-group ISC. As components were extracted from both the narrative-attending and short-stimuli-attending groups, for each to-be-classified participant we obtain a subject-to-group correlation measure with respect to both groups of participants. These measures are from now on referred to as ISC-to-narrative-attending (ISC-NA) and ISC-to-short-stimuli-attending (ISC-SSA). To avoid training biases in the component extraction step, data from the to-be-classified participant were excluded in this step. Paired-sample t-tests were conducted to test whether ISC-NA and ISC-SSA were significantly different within each attentional group. Classification of directed attentional focus was based on the maximum of ISC-NA and ISC-SSA, i.e., if for participant k ISC-NA is larger than ISC-SSA, participant k is classified as a narrative-attending participant. Chance-level classification performance was determined using surrogate data with randomized group labels (narrative-attending/short-stimuli-attending). Significance levels were determined by comparing actual classification with 10000 renditions of randomized group labels.

Autonomic subject-to-group correlations

For each to-be-classified participant, autonomic inter-subject correlations were averaged over all participants in each attentional group to obtain two subject-to-group correlation measures, with respect to the narrative-attending group (ISC-NA) and with respect to the short-stimuli-attending group (ISC-SSA).

Again, paired-sample t-tests were conducted to test whether ISC-NA and ISC-SSA were significantly different within each attentional group. Classification of directed attentional focus was based on the maximum of ISC-NA and ISC-SSA. Significance levels were again determined using surrogate data with randomized group labels.

2.5.6. Effect of interspersed stimuli on neural synchrony

To investigate what effect interspersed stimulus presentation had on PS in EEG in both attentional groups, CorrCA was conducted separately for each of the two groups. This was not only done for EEG responses obtained across the entire audiobook, but also for two subsets of responses, only con-

taining EEG obtained during presentation of WMT trials or IADS-2 trials, respectively. Correlated components were not separately obtained for the subset of potentials obtained during SSST presentation, as the short single trial would not result in enough data to reliably obtain correlated components. Independent-sample t-tests were performed to test for significant differences between attentional groups within each set of responses. Paired-sample t-tests were performed to test for significant differences between EEG sets from either WMT and IADS-2 trials and the set obtained during the entire audiobook within each attentional group. For completeness and comparability with earlier work, these analysis were also conducted separately for each of the three correlated components.

3

Results

3.1. Self-report measures indicative of selective attention

We verified that participants followed their attentional instructions by asking participants to answer questions about the audiobook and short stimuli after the experiment (see Section D.2). Short-stimuli-attending participants could describe significantly more affective sounds (Mdn $= 6$), than narrative-attending participants (Mdn $= 4$), $Z = 2.68$, $p = .007$ and could better estimate the number of tones in the WMT task than narrative-attending participant, with significantly smaller estimation error for short-stimuli attending participants (Mdn $= 1$) than participants attending to the narrative (Mdn $= 10.5$), $Z = 2.82$, $p = .005$. Narrative-attending participants answered more audiobook questions correctly (Mdn $= 6.5$) than short-stimuli-attending participants (Mdn $= 3.5$), $Z = 2.68$, $p = .007$.

3.2. Self-report measures of attention, workload and emotional experience

Directly after the experiment, we asked participants to report their experiences using questions covering attention, workload en emotional experience (see Section D.1). Short-stimuli-attending participants reported to be significantly more stressed during the SSST ($M = 0.54$, $SD = 0.31$) than narrative-attending participants ($M = 0.19$, $SD = 0.22$), $t(24) = 3.35$, $p = .003$. Other than that, experiences did not vary between attentional groups. Short-stimuli-attending (SSA) and narrative-attending (NA) participants reported similar for the amount of distraction during the experiment (SSA: $M = 0.54$, $SD = 0.17$, NA: $M = 0.60$, $SD = 0.26$, $t(24) = 0.63$, $p = .535$), for experienced workload (SSA: $M = 0.61$, $SD = 0.26$, NA: $M = 0.52$, $SD = 0.18$, $t(24) = 1.04$, $p = .312$) and for emotional experience during IADS-2 presentation (SSA: $M = 0.26$, $SD = 0.21$, NA: $M = 0.20$, $SD = 0.17$, $t(24) = 0.71$, $p = .485$).

3.3. Stimulus-locked physiological response traces

EEG event-related potentials, obtained from frontal (Fz), central (Cz) and parietal sites (Pz) distributed along the anterior-posterior midline of the scalp, time-locked to interspersed-stimulus onset are shown

in Figure 3.1. Independent-sample running t-tests, adjusted for multiple comparisons by false discovery rate, reported significant differences between groups in all three midline potentials in response to WMT tone presentation, averaged over all tones in each sequence of tones (FDR adjusted p-values over time are shown in Figure 3.1a). WMT performing participants show greater late positive potentials at Cz and Pz and greater late negative potentials at Fz than participants not fulfilling this task. Neural potentials in response to the first tone of the WMT sequences, shown in Figure F.2, were not significantly different between groups. EEG potentials in response to affective sounds were similar for both attentional groups. Statistical tests only revealed a significant between-group difference for a very short period of time in Cz (see Figure 3.1b). Both attentional groups showed late positive deflections at all three midline sites in response to this stimulus.

Response traces were also obtained for the phasic EDA component and IBI. Figure 3.2 presents these response traces stimulus-locked to onset of WMT, IADS-2, and SSST trials, averaged over participants in each group. Figure 3.2a shows z-scored phasic EDA responses and Figure 3.2b shows IBI responses. Although mean electrodermal responses locked to WMT trials might suggest differently, multiple comparisons adjusted running t-tests did not show significant between group differences (see Figure 3.2a for FDR adjusted p-values). Responses were not different between attentional groups in response to IADS-2 sounds and the SSST either. For IBI, no between-group differences were found either.

3.4. Identifying directed attention using physiological synchrony

To identify the selectively directed attentional state of participants, the similarity of their physiological responses with respect to each of the two groups of individuals with known selectively directed attentional instructions was assessed. Figure 3.3 shows the ISC averaged across participants of the narrative group (left bars) and the short-stimuli group (right bars) when paired with participants of the narrative group (dark bars) or short-stimuli group (light bars). Mean and standard deviation ISC averaged across each group and paired to each group, as well as paired-sample t-test statistics are shown in Table 3.1.

For EEG (Figure 3.3a) ISC is higher for participants when paired to participants of their own attentional group compared to participants from the other group. This is the case for narrative-attending participants ($p = .004$) as well as short-stimuli-attending participants ($p = .004$). For EDA (Figure 3.3b), the same pattern in results is observed, but it only reaches significance for the short-stimuli group ($p = .002$, narrative-attending group: $p = .357$). For IBI (Figure 3.3c), the trend is again the same, but no significant differences were observed (narrative group: $p = .413$; short-stimuli group: $p = .196$).

Assuming for each participants that ISC were higher with participants in the same attentional condition than with those in the other conditions, classification accuracies are significantly higher than chance for EEG and EDA. For EEG, classification accuracy is 85%, both for narrative-attending and short-stimuli-attending participants. For EDA, classification accuracy is 85% for the short-stimuli group and 77% for the narrative group. In IBI, classification accuracy is not significantly higher than chance for both narrative-attending (46%) and short-stimuli-attending participants (61%). Figure 3.4 presents an overview of the classification performance for all measures.

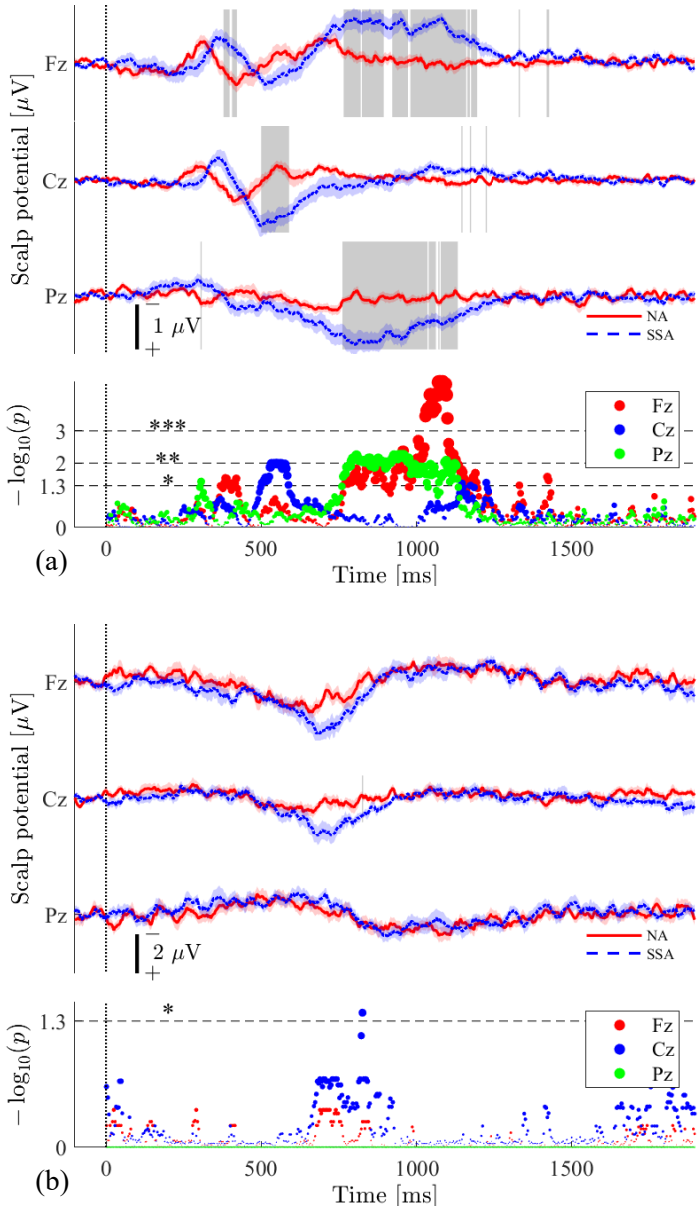


Figure 3.1 Centro-parietal midline event-related potentials time-locked to stimulus onset of the working memory task (a) and affective sounds (b), averaged over narrative-attending participants (red, -) and short-stimuli-attending participants (blue, - -). The standard error of the mean across participants in each group is depicted in shaded areas around the grand average potentials. Significant between group differences ($p < 0.05$, multiple-comparisons adjusted independent-sample t-tests) are depicted with gray areas in the potential plots. Additionally, the corresponding adjusted p-values of the significance tests, including significance thresholds ($*p < .05$, $**p < .01$, $***p < .001$), are shown on a logarithmic scale below the potential plots.

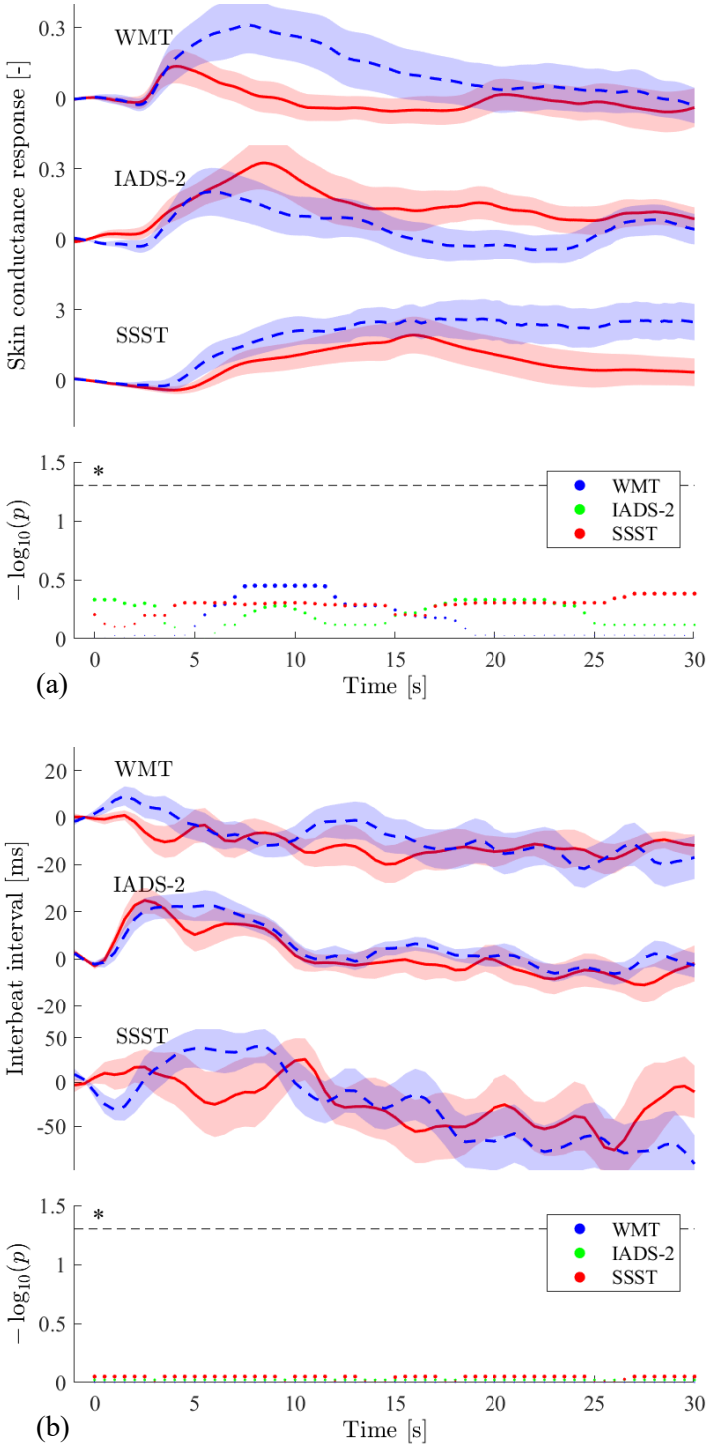
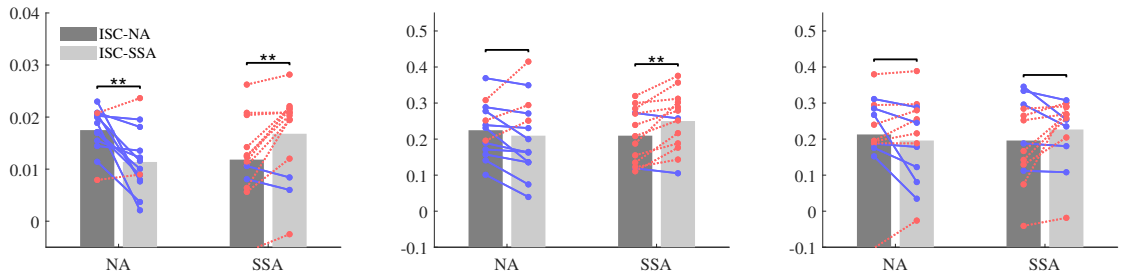


Figure 3.2 Electrodermal (a) and cardiac inter-beat interval (b) responses time-locked to stimulus onset of the working memory task (WMT), affective sounds (IADS-2) and sing-a-song stress test (SSST), averaged over narrative-attending participants (red, -) and short-stimuli-attending participants (blue, - -). The standard error of the mean across participants in each group is depicted in shaded areas around the average response traces. No significant between-group differences ($p < 0.05$, multiple-comparisons adjusted independent-sample t-tests) were found. The adjusted p-values of the significance tests, including significance threshold ($*p < .05$), are shown on a logarithmic scale below the response traces.



(a) Electroencephalography **(b)** Electrodermal activity **(c)** Inter-beat interval

Figure 3.3 Subject-to-group physiological inter-subject correlations with respect to the narrative-attending group(ISC-NA) and short-stimuli-attending group (ISC-SSA) for narrative-attending participants (NA) and short-stimuli-attending participants (SSA) for (a) EDA, (b) IBI and (c) EEG. Connected dots display subject-to-group correlations of one individual to both groups, where blue (–) lines display individuals for which ISC-NA > ISC-SSA and red (···) lines display individuals for which ISC-SSA > ISC-NA. Paired-sample t-tests revealed within-group correlations were higher than between-group correlations in EEG and EDA (test statistics are presented in Table 3.1) ($^{**}p < .01$).

3.5. Effect of stimulus-presentation on inter-subject correlations

Analysis were then focused on EEG. Figure 3.5 shows ISC across narrative-attending (NA) and short-stimuli-attending (SSA) participants for the set of potentials obtained during the entire naturalistic stimulus (All) and for two subsets of responses, only containing potentials obtained during presentation of WMT and IADS-2 trials. Gray bars depict chance-level correlations, obtained using phase-randomized surrogate data. ISC computed as the sum of correlations in the first three correlated components were much higher than chance for each (sub)set of potentials independent of attentional condition. Test statistics for this analysis are presented in Section F.4. Here, significance with respect to chance level is also tested separately for each of the three components.

When considering the duration of the entire audiobook, narrative-attending and short-stimuli-attending participants showed similar amount of ISC, not resulting in a significant difference ($p = .698$). Also when considering only potentials obtained during affective sound presentation, ISC were not significantly different between groups ($p = .528$). However, when considering the subset of potentials obtained during WMT trials, responses were shown to be more correlated accross short-stimuli-attending participants then narrative-attending participants ($p = .033$). Test statistics are presented in Table 3.2.

For each attentional group, ISC values of potentials obtained during short-stimuli presentation were compared to those of potentials from the entire narrative. Test statistics are shown in Table 3.3. ISC

Table 3.1 Mean (SD) subject-to-group inter-subject correlations of narrative-attending participants (NA) and short-stimuli-attending participants (SSA), paired with the narrative group (ISC-NA) and short-stimuli group (ISC-SSA), for EEG, EDA and IBI. In addition, test statistics of paired-sample t-tests between ISC-NA and ISC-SSA in each group are presented.

	NA		SSA	
	ISC-NA	ISC-SSA	ISC-NA	ISC-SSA
EEG	.018 (.004)	.012 (.006)	.017 (.008)	.011 (.008)
	$t(12) = 3.57, p = .004$		$t(12) = -3.57, p = .004$	
EDA	.225 (.075)	.210 (.107)	.210 (.076)	.250 (.081)
	$t(12) = 0.96, p = .357$		$t(12) = -3.92, p = .002$	
IBI	.213 (.115)	.196 (.117)	.196 (.113)	.227 (.092)
	$t(12) = 0.85, p = .413$		$t(12) = -1.37, p = .196$	

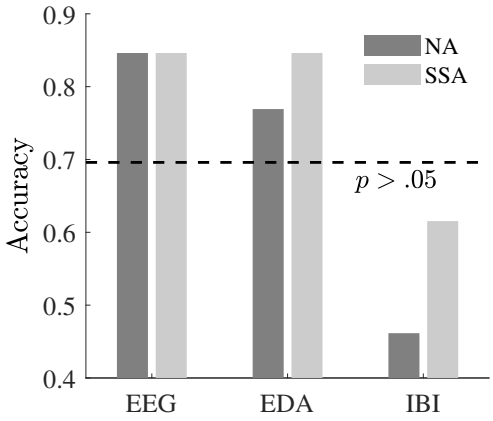


Figure 3.4 Selectively directed attention classification accuracy using inter-subject correlations in electroencephalography (EEG), electrodermal activity (EDA) and inter-beat interval (IBI) over narrative-attending (NA) and short-stimuli-attending (SSA) participants. Classification accuracy below the dashed line is not significantly higher than chance level ($p > .05$).

of the short-stimuli group were significantly higher during short-stimulus presentation than during the entire narrative, for both WMT ($p < .001$) and IADS-2 trials ($p = .001$). ISC of the narrative-attending group were only significantly higher compared to the entire narrative during IADS-2 presentation ($p = .003$), not during WMT presentation ($p = .193$).

For the sake of completeness and comparability to earlier work, the above-presented analyses were also conducted for each of the first three correlated components separately. This analysis is presented in Section F.3. The scalp distribution of the correlated components during the entire narrative are also presented there, for both narrative-attending and short-stimuli-attending participants (Figure F.3).

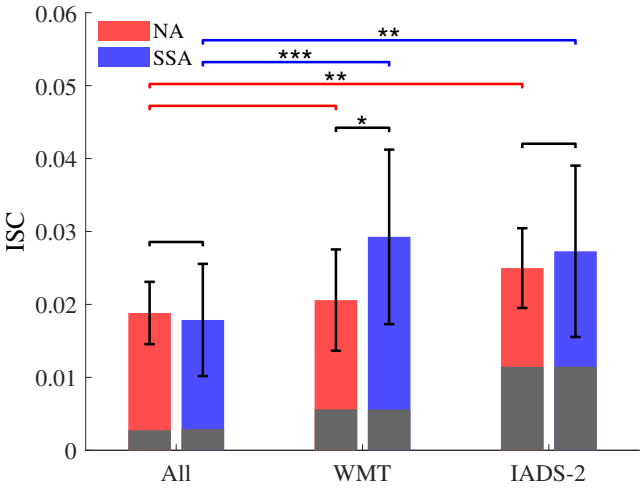


Figure 3.5 Inter-subject correlations (ISC) between EEG responses of narrative-attending participants (NA, red) and short-stimuli-attending participants (SSA, blue), computed across the entire narrative-stimulus (All), and subsets of potentials corresponding to presentation of the working memory task (WMT) and affective sounds (IADS-2). Gray bars depict chance level correlations. Independent sample t-tests revealed higher ISC in the short-stimuli-attending group than the narrative-attending group during WMT (see Table 3.2). Paired sample t-tests revealed higher ISC during WMT and IADS-2 presentation compared to the complete narrative for group SSA (blue significance bars) and only during IADS-2 presentation for group NA (blue significance bars) (see Table 3.3) (*$p < .05$, **$p < .01$, ***$p < .001$).

Table 3.2 Independent-sample t-test statistics of inter-subject correlation differences between narrative-attending and short-stimuli-attending participants for the set of potentials obtained during the entire narrative stimulus (All) and for subsets only containing potentials obtained during working memory task (WMT) or affective sound (IADS-2) presentation.

All	WMT	IADS-2
$t(24) = -0.39, p = .698$	$t(24) = 2.26, p = .033$	$t(24) = 0.64, p = .528$

Table 3.3 Paired-sample t-test statistics of between-stimulus-condition inter-subject correlation differences in narrative-attending (NA) and short-stimuli-attending (SSA) participants.

	NA	SSA
WMT vs. All	$t(12) = 1.38, p = .193$	$t(12) = 8.11, p < .001$
IADS-2 vs. All	$t(12) = 3.77, p = .003$	$t(12) = 4.17, p = .001$

4

Discussion and conclusions

This study was aimed at examining whether PS in autonomic and neural measures are suitable markers to recover selectively directed attentional focus and moments of emotional or task-related relevance. In a between-subjects experiment, EEG, EDA and IBI were monitored from participants who were instructed to either attend to a narrative audiobook ($n = 13$) or to interspersed auditory events ($n = 13$), that were emotional sounds, tones that attending participants needed to keep track of and a stress test attending participants needed to perform.

4.1. Stimulus-locked physiological response traces

To obtain an indication of the suitability of each of the physiological measures in distinguishing between different attentional instructions, it was investigated whether deflections of EEG, EDA and IBI in response to interspersed stimuli were larger for the short-stimuli attending group than for the narrative-attending group. Deflections in physiological responses toward interspersed stimuli could partly distinguish between attentional groups. Only EEG signals in response to sequences of tones that attending participants had to keep track of (working memory task - WMT), were larger for task task-attending participants than book-attending participants. EDA and IBI responses were not different between groups. In response to affective sounds or the sing-a-song stress test attending participants had to perform, physiological responses were similar between groups.

It might have been expected that WMT better distinguished between attentional groups than affective sound trials. The WMT task largely attracts attention by 'top-down' task instructions, that were not provided to book-attending participants. This is unlike affective sounds, that mainly attract attention through 'bottom-up' mechanisms of emotional relevance or salience. That responses during the SSST were not different between groups was not expected and did not correspond with self-reported stress, which was significantly higher for the short-stimuli-attending group than for the narrative-attending group.

That EEG performed best in distinguishing attention in the WMT is not unexpected based on earlier work. EEG event-related potentials, especially at Pz, more reliably predict workload than EDA or HR

(Hogervorst et al., 2014) and are known to be directly reflective of attentional and working memory processes (Brouwer et al., 2012; Polich, 2007; Polich and Kok, 1995).

4.2. Identifying directed attention using physiological synchrony

It was then examined for each of the physiological measures whether responses were more synchronized between participants in the same attentional condition than between participants with different attentional instructions. PS in both EEG and EDA were indicative of shared attention. EEG and EDA signals of participants were more strongly synchronized with those of participants in the same attentional condition than with those of participants in the other attentional condition. For IBI, we did not find such results. The directed attentional focus of each participant was classified as indicated by the group with whom the participant showed the highest averaged PS. EEG data allowed attribution to the correct attentional group in 85% of the cases, for EDA this was 81% and for IBI accuracy was below chance level. Both autonomic (EDA) and neural (EEG) measures were thus found to be suitable markers for the recovery of directed attention.

That EEG PS performed relatively well might have been expected based on previous PS literature, the more direct link with attention and the results discussed above. That EDA did well too is convenient from a user perspective, as EDA can be more easily monitored in real-world settings using one of the many available wearable systems (e.g., Asada et al., 2003; Garbarino et al., 2014; Poh et al., 2010). The finding that IBI performed worst may not be unexpected given the fact that the relation between heart rate and mental state seems less straightforward than EDA. Whereas EDA has consistently been found to be positively related to arousal (Andreassi, 2010), the relation between emotional stimuli and heart rate has been found to be more complex. Both positive (e.g., Brouwer and Hogervorst, 2014) and negative (e.g., Brouwer et al., 2015) relations between heart rate and arousal have been reported. The reason for this is probably that arousal can be associated with the body being prepared for action, the defense reflex, or with a concentrated, focused state, the orienting reflex. The defense system is associated with heart rate accelerations and the orienting system with decelerations (Graham and Clifton, 1966).

4.3. Inter-subject correlations of physiological responses

Last, the effect of interspersed stimulus presentation on PS in EEG of both attentional groups was examined, aiming to explore the potential of using PS for the detection of moments of emotional or task-related relevance. These analysis were done using EEG, since stimulus-locked EEG potentials best distinguished between groups and PS in EEG best predicted attentional state of participants.

PS was found to be higher during presentation of short stimuli than over the entire experiment. For the SSA group this was the case for WMT and IADS-2 trials, and for the NA group only during IADS-2 trials. This difference between groups can be attributed to difference in relevance of the events between groups. WMT trials were only relevant for the SSA group, as only they received task instructions. The IADS-2 sounds on the other hand, had emotional relevance for both groups and thus automatically drew attention of all participants. These results suggest PS across individuals with shared attention might be a promising marker of the occurrence of moments of emotional or task-related relevance.

4.4. Limitations and recommendations

Some limitations of the current work and recommendations for future work must be noted. In the current study, autonomic PS was computed using simple Pearson correlations in a moving window. Not only is this method computationally inexpensive and easily adaptable for online use, it also provides inter-subject correlations over time, making it suitable for eventual detection of emotionally or task-related relevant events in time. However, this method has its limitations, including oversampling as a result of overlapping windows as well as potentially spurious correlations as a result of not controlling for autocorrelation (Levenson and Gottman, 1983). Future research should investigate whether other methods would result in the similar findings as obtained in this study. This could start with varying window sizes and step increments of the current method, but many other available approaches for PS assessment could also be tested, such as dynamical correlation (Liu et al., 2016) or wavelet analysis (Quer et al., 2016).

The potential of PS as a marker of the occurrence of moments of emotional or task-related relevance was explored using EEG only. This does not mean that EDA or other measures cannot be suited for relevant-event detection. Especially the link with emotional relevance is well established for EDA (Bradley et al., 1996; Bradley and Lang, 2007; Brouwer and Hogervorst, 2014). Future studies should therefore look into the potential of EDA for detection of moments of relevance. Future analysis could even examine patterns of synchrony in the different modalities as a function of stimulus type. Events relevant for mental tasks (WMT) may be relatively strongly associated to synchrony in EEG, whereas emotional stimuli (IADS-2 and the instruction to sing a song) may be relatively strongly associated to synchrony in EDA. Patterns of multimodal synchrony might even allow us to identify the source of shared attention, or in other words, the type of shared mental activity.

Another note is on the method of neural PS computation during presentation of short stimuli. For examination of the effect of stimulus presentation on PS in each group, correlated components were separately extracted for each attentional group and for each (sub)set of data with potentials during trials of the specific stimuli. The resulting components thus capture the most correlated brain activity during trials of the specific short-stimulus type only. One could also extract components across the entire narrative stimulus and only separately project the different (sub)sets of data on these common extracted components vectors. This approach would be a next step toward identification of emotionally or cognitively relevant events, as the component extraction step is then no longer dependent on known event markers.

4.5. Conclusions

In conclusion, PS in EEG and EDA are suitable markers of selectively directed attention. PS was higher between participants with shared attentional instruction than participants with opposing attentional instruction and allowed attribution to the correct attentional group for 85% of the participants for EEG and 81% for EDA. PS across shared-attending individuals is also promising for the detection of moments of emotional or task-related relevance. EEG PS was higher across participants during presentation of relevant interspersed stimuli than over the entire experiment. The findings of this study are a first step toward identification of directed attention and emotional or task-related relevance in real-world environments and contribute to increased understanding of physiological mechanisms of attention.

Bibliography

Andreassi, J. L. (2010). *Psychophysiology: Human behavior and physiological response*. Psychology Press. doi:10.4324/9780203880340.

Asada, H. H., Shaltis, P., Reisner, A., Rhee, S., and Hutchinson, R. C. (2003). Mobile monitoring with wearable photoplethysmographic biosensors. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):28–40. doi:10.1109/MEMB.2003.1213624.

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159. doi:10.1162/neco.1995.7.6.1129.

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, 22(3):292–299. doi:10.1111/j.1469-8986.1985.tb01603.x.

Benedek, M. and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1):80–91. doi:10.1016/j.jneumeth.2010.04.028.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of erp components — a tutorial. *NeuroImage*, 56(2):814–825. doi:10.1016/j.neuroimage.2010.06.048.

Borovac, A., Stuldreher, I. V., Thammasan, N., and Brouwer, A.-M. (submitted). Towards measuring group physiological synchrony using wearables: validation of wearables for electrodermal activity and heart rate.

Bradley, M. M., Cuthbert, B. N., and Lang, P. J. (1996). Picture media and emotion: Effects of a sustained affective context. *Psychophysiology*, 33(6):662–670.

Bradley, M. M. and Lang, P. J. (2007). The International Affective Digitized Sounds (IADS-2): Affective ratings of sounds and instruction manual. *University of Florida, Gainesville, FL, Tech. Rep. B-3*.

Brouwer, A.-M., Hogervorst, M., Reuderink, B., van der Werf, Y., and van Erp, J. (2015). Physiological signals distinguish between reading emotional and non-emotional sections in a novel. *Brain-computer interfaces*, 2(2-3):76–89.

Brouwer, A.-M. and Hogervorst, M. A. (2014). A new paradigm to induce mental stress: the Sing-a-Song Stress Test (SSST). *Frontiers in Neuroscience*, 8:224. doi:10.3389/fnins.2014.00224.

Brouwer, A.-M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, 9(4):045008. doi:10.1088/1741-2560/9/4/045008.

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *The Neuroscientist*, 8(2):132–142. doi:10.1177/107385840200800209.

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 without secondary tasks. *Multisensory Research*, 31(3-4):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. *Journal of Neuroscience Methods*, 134(1):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. *Nature Communications*, 5:4567. doi:10.1038/ncomms5567.

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? *Frontiers in Human Neuroscience*, 6:112. doi:10.3389/fnhum.2012.00112.

Elkins, A. N., Muth, E. R., Hoover, A. W., Walker, A. D., Carpenter, T. L., and Switzer, F. S. (2009). Physiological compliance and team performance. *Applied Ergonomics*, 40(6):997–1003. doi:10.1016/j.apergo.2009.02.002.

Garbarino, M., Lai, M., Bender, D., Picard, R. W., and Tognetti, S. (2014). Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, pages 39–42. IEEE. doi:10.4108/icst.mobihealth.2014.257418.

Graham, F. K. and Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological bulletin*, 65(5):305. doi:10.1037/h0023258.

Hanson, S. J., Gagliardi, A., and Hanson, C. (2009). Solving the brain synchrony eigenvalue problem: conservation of temporal dynamics (fMRI) over subjects doing the same task. *Journal of Computational Neuroscience*, 27(1):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 in Cognitive Sciences*, 14(1):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(5664):1634–1640. doi:10.1126/science.1089506.

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. *Frontiers in Neuroscience*, 8:322. doi:10.3389/fnins.2014.00322.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28(3/4):321–377. doi:10.2307/2333955.

Ki, J. J., Kelly, S. P., and Parra, L. C. (2016). Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *Journal of Neuroscience*, 36(10):3092–3101. doi:10.1523/jneurosci.2942-15.2016.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, T. A. Williams (Eds). *Technology in Mental Health Care Delivery Systems*.

Levenson, R. W. and Gottman, J. M. (1983). Marital interaction: physiological linkage and affective exchange. *Journal of Personality and Social Psychology*, 45(3):587. doi:10.1037//0022-3514.45.3.587.

Liu, S., Zhou, Y., Palumbo, R., and Wang, J.-L. (2016). Dynamical correlation: A new method for quantifying synchrony with multivariate intensive longitudinal data. *Psychological Methods*, 21(3):291. doi:10.1037/met0000071.

Marci, C. D. (2006). A biologically based measure of emotional engagement: Context matters. *Journal of Advertising Research*, 46(4):381–387. doi:10.2501/s0021849906060466.

Marci, C. D., Ham, J., Moran, E., and Orr, S. P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *The Journal of Nervous and Mental Disease*, 195(2):103–111. doi:10.1097/01.nmd.0000253731.71025.fc.

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. *Personality and Social Psychology Review*, 21(2):99–141. doi:10.1177/1088868316628405.

Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage*, 28(2):326–341. doi:10.1016/j.neuroimage.2005.05.032.

Percival, D. B. and Walden, A. T. (2006). *Wavelet methods for time series analysis*, volume 4. Cambridge university press.

Poh, M.-Z., Swenson, N. C., and Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical engineering*, 57(5):1243–1252. doi:10.1109/tbme.2009.2038487.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10):2128–2148. doi:10.1016/j.clinph.2007.04.019.

Polich, J. and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology*, 41(2):103–146. doi:10.1016/0301-0511(95)05130-9.

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. *Scientific Reports*, 7:43916. doi:10.1038/srep43916.

Quer, G., Daftari, J., and Rao, R. R. (2016). Heart rate wavelet coherence analysis to investigate group entrainment. *Pervasive and Mobile Computing*, 28:21–34. doi:10.1016/j.pmcj.2015.09.008.

Stuldreher, I. V., De Winter, J. C. F., Thamassan, N., and Brouwer, A.-M. (submitted). Analytic approaches for the combination of autonomic and neural activity in assessment of physiological synchrony.

The Mathworks, Inc. (2015). R Wave Detection in the ECG. https://nl.mathworks.com/help/wavelet/ug/r-wave-detection-in-the-ecg.html. Accessed: 19 March 2019.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1-4):77–94. doi:10.1007/s00704-005-0130-7.

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 *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4101–4105. IEEE. doi:10.1109/embc.2015.7319296.

Winkler, I., Haufe, S., and Tangermann, M. (2011). Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behavioral and Brain Functions*, 7(1):30. doi:10.1186/1744-9081-7-30.

World Medical Association (2014). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists*, 81(3):14.

List of Figures

List of Tables

A

Information for participants

This appendix presents the forms 'study information for participants' (Informatie voor deelnemers aan niet WMO-plichtig onderzoek), 'informed consent' (toestemmingsverklaring) and group-specific task instructions for the narrative-attending group (Groep 1) and the short-stimuli-attending group (Groep 2), which were provided to participants before participating in this study. Note that these forms are only presented in Dutch.

Informatie voor deelnemers aan <u>niet WMO-plichtig</u> onderzoek

---

Titel:	Monitoring Mental State using Physiological Synchrony
Projectnummer:	060.29994
Onderzoeksleider: Telefoonnummer:	Anne-Marie Brouwer +31 6 46847276 anne-marie.brouwer@tno.nl
Onderzoekslocatie:	**TNO Soesterberg** Kampweg 55 3769 DE Soesterberg
Adres:	**TNO Soesterberg** Postbus 23 3769 ZG Soesterberg 088 866 15 00

1 Wat is het doel van het onderzoek?

Wij onderzoeken methoden om mentale staat, zoals aandacht en stress, op een continue manier te monitoren, zonder herhaaldelijk vragen te hoeven stellen. Hiervoor richten wij ons met name op het gebruik van fysiologische metingen, zoals hartslag, zweetrespons en hersensignalen.

In dit onderzoek willen we ontdekken of, en op welke manier, we kunnen inschatten waar de aandacht van een persoon op is gericht. We vergelijken hierbij ook laboratorium apparatuur met de meer draagbare en goedkope varianten. Uiteindelijk kan dit onderzoek van nut zijn om personen te ondersteunen in het richten van aandacht, of om te onderzoeken welke elementen in een omgeving aandacht trekken.

2 Over TNO

De letters TNO staan voor Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek. TNO ontwikkelt kennis gericht op praktische toepassing en richt zich hierbij op de volgende aandachtsgebieden: Bouw, Infra & Maritiem; Circulaire Economie & Omgeving; Defensie & Veiligheid; Energie; Gezond Leven; Industrie; Informatie & Communicatie Technologie; Mobiliteit & Logistiek en tot slot Strategische Analyses & Beleid.

3 Deelname aan onderzoek

Via dit document willen wij u informeren over het onderzoek waaraan u wordt gevraagd deel te nemen. Het onderzoek zal plaats vinden bij TNO Soesterberg, waarvan het bezoekadres op de voorpagina is vermeld. Het onderzoek zal inclusief instructie maximaal twee uur in beslag nemen. In totaal zullen we ongeveer 25 deelnemers onderzoeken voor deze studie.

4 Wie kan meedoen aan het onderzoek?

Selectiecriteria zijn een goed gehoor en geen u bekende problemen met aandacht (u mag bijvoorbeeld niet lijden aan ADHD of ADD).

5 Hoe wordt het onderzoek uitgevoerd?

Het onderzoek zelf duurt ongeveer één uur (inclusief voorbereiding, instructie en nabespreking, zal het maximaal twee uur in beslag nemen), en vindt plaats bij TNO Soesterberg. Tijdens het onderzoek zullen verschillende fysiologische maten worden gemeten. U krijgt sensors op uw vingers om uw zweetrespons te meten (electrodermal activity), op uw borst om uw hartslag te meten (electrocardiography of ECG) en op uw hoofd om signalen in het brein te meten (electroencephalography of EEG). Daarnaast zult u ook nog twee draagbare hartslagmeters in de vorm van een borstband en een polsband omkrijgen, en een draagbare zweetrespons sensor op uw handpalm. Gedurende het onderzoek krijgt u audio te horen die is samengesteld uit een luisterboek en andere geluiden. Afhankelijk van de groep waarin u bent ingedeeld vragen we u deze geluiden te negeren of juist niet.

6 Wat wordt er van u verwacht?

Wij vragen u, afhankelijk van de groep waarin u bent ingedeeld, ofwel alleen te focussen op het luisterboek en de geluiden te negeren of u juist te richten op de geluid en eventueel bijbehorende mentale taakjes uit te voeren.

Daarnaast vragen wij u om tijdens het onderzoek stil te blijven zitten, zodat de fysiologische metingen niet worden beinvloed door bewegingen.

7 Wat zijn mogelijk voor- en nadelen van deelname aan dit onderzoek?

Voor het meten van hartslag is het nodig sensoren te plakken bij het rechtersleutelbeen en ter hoogte van de onderste ribben; de hartslag band komt wat hoger direct op de huid. Hiervoor is het nodig kleding tijdelijk opzij te doen. Voor het meten van de brein activiteit wordt gebruik gemaakt van gel die in het haar kan achterblijven. Er wordt de mogelijkheid geboden om na het onderzoek de haren te wassen. Daarnaast kan er wat verhoogde stress plaatsvinden door de mentale taken.

Door deel te nemen aan dit onderzoek helpt u mee met het ontwikkelen van kennis. Zo hopen wij modellen die fysiologische metingen gebruiken om de mentale staat van een persoon in te schatten te verbeteren. Daarnaast willen we in dit onderzoek de bruikbaarheid testen van zogenaamde wearables ten opzichte van labaratorium apparatuur.

8 Wat gebeurt er als u niet (meer) wenst deel te nemen aan dit onderzoek?

Deelname is geheel vrijwillig en de deelnemer kan op elk moment stoppen met het onderzoek zonder opgave van reden en verdere consequenties. Indien van toepassing kan het onderzoek ook op elke moment gestaakt worden door de onderzoeker.

9 Wat gebeurt er met uw gegevens?

Tijdens dit onderzoek worden de volgende gegevens van u verzameld:

- Zweetresponds (Electrodermal activity)
- Hartslag (Electrocardiography)
- Brein activiteit (Electroencephalography)
- Antwoorden op de vragenlijsten
- Uw leeftijd en geslacht

Uw persoonsgegevens worden gepseudonimiseerd. Dat houdt in dat uw data wordt gekoppeld aan een proefpersoonnummer. Fysiologische data wordt ook alleen onder dit proefpersoonnummer opgeslagen.

TNO hecht groot belang aan uw privacy neemt de daarvoor geldende regels in acht. Naam- en adresgegevens van deelnemers worden direct na de verzending van de beloningen vernietigd. Om uw gegevens te beveiligen, zijn tal van maatregelen getroffen. De gegevens worden verwerkt met goed beveiligde computersystemen waartoe onbevoegden geen toegang hebben. TNO garandeert dat uw gegevens alleen voor statistische doeleinden worden gebruikt. Derden hebben geen toegang tot de verzamelde gegevens. In publicaties over het onderzoek zijn de (antwoorden van) individuele deelnemers op geen enkele wijze herkenbaar.

Uw gegevens worden zorgvuldig behandeld. Om uw privacy te waarborgen, worden uw naam en contactgegevens zoveel mogelijk gescheiden van uw onderzoeksgegevens bewaard. Dit bevordert uw anonimiteit, zelfs in het zeer onwenselijke en onwaarschijnlijke geval dat onderzoeksgegevens in handen komen van onbevoegden door bijvoorbeeld verlies, diefstal, misbruik of onbevoegde toegang. Uw gegevens zijn slechts toegankelijk voor daartoe bevoegde leden van het onderzoeksteam. Inzage door bevoegde inspecteurs kan nodig zijn om de betrouwbaarheid en kwaliteit van het onderzoek na te gaan. Na afloop van het onderzoek kunnen uw onderzoeksgegevens gedurende 15 jaar na afloop van het onderzoek worden bewaard.

10 Is er een vergoeding wanneer u besluit aan dit onderzoek mee te doen?

Voor deelname aan dit onderzoek krijgt u een financiële vergoeding van €30. Daarnaast krijgt u een reiskostenvergoeding van 19 cent/km gerekend vanaf het woonadres tot TNO Soesterberg, met een max. van €15,- per persoon.

TNO is verplicht de aan u betaalde vergoeding voor deelname op te geven aan de Belastingdienst.

11 Ethische aspecten

TNO gaat zorgvuldig met u om. U doet vrijwillig mee en u krijgt precies te horen wat u moet doen. Als u het daar mee eens bent en u bent geschikt om mee te doen dan begint u aan het onderzoek. U kunt ook stoppen gedurende het onderzoek als het u niet (meer) bevalt. U hoeft daarbij geen reden op te geven. Het onderzoek wordt uitgevoerd volgens alle van toepassing zijnde nationale en internationale wetgeving en richtlijnen die gericht zijn op het bewaken van uw gezondheid en veiligheid.

12 Verzekering

Voor iedereen die meedoet aan dit onderzoek heeft TNO een verzekering afgesloten. De verzekering dekt schade door deelname aan het onderzoek. Schade moet u zo snel mogelijk aan TNO melden.

13 Wilt u verder nog iets weten?

Indien u nog vragen heeft of meer informatie wilt hebben kunt u contact opnement met de proefleider Ivo Stuldreher via ivo.stuldreher@tno.nl of de projectleider Anne-Marie Brouwer via anne-marie.brouwer@tno.nl.

14 Ondertekening toestemmingsformulier

In de bijlage vindt u de toestemmingsverklaring. Wanneer u geen vragen meer heeft en deel wilt nemen aan dit onderzoek, vragen we u deze in te vullen.

Informed consent / toestemmingsverklaring

Ondergetekende,

Naam . _____

Geboortedatum _____

verklaart op vrijwillige basis deel te nemen aan het onderzoek, getiteld

'Monitoring Mental State using Physiological Synchrony'

_____

bij TNO.

Ik bevestig dat ik de informatie over bovengenoemd onderzoek heb gelezen.
Ik begrijp de informatie.

De bedoelingen van het onderzoek en de daarbij gevolgde aanpak zijn tot mijn tevredenheid uitgelegd.

Ik heb de gelegenheid gehad om aanvullende vragen te stellen en deze vragen zijn naar tevredenheid beantwoord.

Ik heb voldoende tijd gehad om over deelname na te denken.

Ik weet dat mijn deelname aan het onderzoek geheel vrijwillig is en dat ik mijn toestemming op ieder moment kan intrekken zonder dat ik daarvoor een reden hoef op te geven.

Ik geef toestemming mijn persoonsgegevens te verwerken voor de doelen zoals beschreven in de informatie.

Ik geef toestemming mijn fysiologische gegevens te verwerken voor de doelen zoals beschreven in de informatie.

Ik verklaar te beschikken over een goed gehoor en verklaar geen bekende aandachtsproblemen te hebben.

Ik geef toestemming voor het bewaren van de gegevens en dat bevoegde leden van het onderzoeksteam en bevoegde inspecteurs hier inzage in hebben.

Ik verklaar me te houden aan de instructies van de proefleider.

Ik ben ermee bekend dat de proefleider de deelname aan het onderzoek kan beëindigen als hij/zij dat nodig vindt.

Voorts verklaar ik geen mij bekende belemmeringen te hebben om aan het onderzoek 'Monitoring Mental State using Physiological Synchrony' deel te nemen.

Plaats, datum _____

Handtekening proefpersoon: _____

TOELATING
Ik heb me ervan vergewist dat ik deze proefpersoon goed geïnformeerd heb over het onderzoek waaraan hij/zij gaat deelnemen. Ik heb mij ervan overtuigd dat deze proefpersoon voldoet aan de selectiecriteria om aan bovengenoemd onderzoek deel te mogen nemen.

Naam, handtekening en datum ondertekening proefleider:

Naam proefleider:	
Datum en plaats:	

Handtekening	

Informatie Groep 1

U krijgt zodadelijk audio te horen die is samengesteld uit een luisterboek en andere geluiden, zoals korte geluidsfragmenten, piepjes en gesproken opdrachten. Deze samengestelde audiostream duurt ongeveer 65 minuten. Tijdens deze audiostream meten we fysiologische maten met de verschillende sensors.

We meten uw fysiologische maten ook voor en na de samengestelde audiostream tijdens twee baseline condities: een stille periode en een periode waar u ruis te horen krijgt. Aan het begin van het onderzoek is het eerst één minuut stil en aansluitend hoort u één minuut ruis. Na het luisterboek is het omgedraaid, u hoort dan eerst één minuut ruis en daarna is het nog één minuut stil.

Om de metingen zo zuiver mogelijk te houden, vragen wij u tijdens het hele onderzoek stil te blijven zitten, niet te praten, uw ogen open te houden en de instructies van de proefleider op te volgen. Ook als het luisterboek stopt vragen we u nog stil te blijven zitten en niet te praten omdat dan de baseline condities nog volgen. De proefleider zal aangeven wanneer de baseline condities voorbij zijn en het experiment is afgerond. De proefleider is altijd op gehoorsafstand in de buurt.

Na het luisteren naar de samengestelde audio vragen wij u nog enkele vragen te beantwoorden, waaronder inhoudelijke vragen over de audio die u te horen heeft gekregen.

Wat wordt er van u verwacht?

U heeft de opdracht om *volledig te focussen op het luisterboek*. Let dus zo goed mogelijk op het boek en negeer alle andere geluiden. Eventuele gesproken opdrachten moet u niet uit te voeren. Achteraf krijgt u vragen over de inhoud van het luisterboek. Probeer deze zo goed mogelijk te beantwoorden.

Tijd (h:mm)	Fase
0:00	Ontvangst
0:05	Instructie
0:10	
0:15	Voorbereiding experiment
0:20	
0:25	
0:30	
0:35	
0:40	Stil + Ruis
0:45	Experiment
0:50	
0:55	
1:00	
1:05	
1:10	
1:15	
1:20	
1:25	
1:30	
1:35	
1:40	
1:45	Ruis + Stil
1:50	Vragen
1:55	
2:00	Opfrissen

Informatie Groep 2

Hoe wordt het onderzoek uitgevoerd?

U krijgt zodadelijk audio te horen die is samengesteld uit een luisterboek en andere geluiden, zoals korte geluidsfragmenten, piepjes en gesproken opdrachten. Deze samengestelde audiostream duurt ongeveer 65 minuten. Tijdens deze audiostream meten we fysiologische maten met de verschillende sensors.

We meten uw fysiologische maten ook voor en na de samengestelde audiostream tijdens twee baseline condities: een stille periode en een periode waar u ruis te horen krijgt. Aan het begin van het onderzoek is het eerst één minuut stil en aansluitend hoort u één minuut ruis. Na het luisterboek is het omgedraaid, u hoort dan eerst één minuut ruis en daarna is het nog één minuut stil.

Om de metingen zo zuiver mogelijk te houden, vragen wij u tijdens het hele onderzoek stil te blijven zitten, niet te praten, uw ogen open te houden en de instructies van de proefleider op te volgen. Ook als het luisterboek stopt vragen we u nog stil te blijven zitten en niet te praten omdat dan de baseline condities nog volgen. De proefleider zal aangeven wanneer de baseline condities voorbij zijn en het experiment is afgerond. De proefleider is altijd op gehoorsafstand in de buurt.

Na het luisteren naar de samengestelde audio vragen wij u nog enkele vragen te beantwoorden, waaronder inhoudelijke vragen over de audio die u te horen heeft gekregen.

Wat wordt er van u verwacht?

U hoeft niet actief te luisteren naar het luisterboek, maar wij vragen u te *focussen op de andere geluiden* die door het luisterboek heen worden afgespeeld. U krijgt gedurende het luisterboek drie verschillende soorten geluiden te horen.

- Gedurende het hele luisterboek krijgt u korte geluidsfragmenten te horen. Luister goed naar deze geluiden.
- Gedurende het hele luisterboek hoort u ook fragmenten met twee verschillende soorten piepjes, een hoge en een lage. Bij elk van deze fragmenten heeft u de opdracht om te tellen hoe vaak u beide piepjes hoort. Spreek het antwoord niet uit, maar zorg wel dat u de piepjes echt telt. Ook als u de tel bent kwijt geraakt is het belangrijk zo goed mogelijk door te gaan. U heeft zometeen de mogelijkheid deze opdracht een keer te oefenen.
- Aan het einde van de hele audiostream na ongeveer een uur, krijgt u ook nog een gesproken opdracht te horen. Voer deze opdracht uit.

Achteraf krijgt u vragen over de geluidsfragmenten en de piepjes.

Tijd (h:mm)	Fase
0:00	Ontvangst
0:05	Instructie
0:10	
0:15	Voorbereiding experiment
0:20	
0:25	
0:30	
0:35	
0:40	Stil + Ruis
0:45	Experiment
0:50	
0:55	
1:00	
1:05	
1:10	
1:15	
1:20	
1:25	
1:30	
1:35	
1:40	
1:45	Ruis + Stil
1:50	Vragen
1:55	
2:00	Opfrissen

B

Literature Study

In advance to this thesis, a literature study was conducted, aimed at exploring analytic approaches for 'A-N' multimodal physiological synchrony assessment. The scope of A-N multimodal synchrony assessment is to use a combination of neural and autonomic measures, for example EDA and EEG, to assess inter-subject physiological synchrony. To enable such combinations, I reviewed methods already used to combine multiple channels from either autonomic or neural measures into a single measure of physiological synchrony, such as the combination of multiple EEG channels into a measure of overall brain-to-brain synchrony.

The paper presented in this Appendix, named 'Analytic approaches for the combination of autonomic and neural activity in the assessment of physiological synchrony', has already been graded for the 'ME-BMD Literature Report (ME51010)' course. The paper as presented here has also been accepted to be presented at the 'Multimodal Brain/Body-Machine Interfaces for 'In-the-Wild' Experiments' special session of the IEEE SMC 2019 conference, that will be held in Bari in October this year. Co-authors of this paper are Joost de Winter, Nattapong Thammasan and Anne-Marie Brouwer, who all carefully reviewed drafts of the paper, for which I am very grateful.

Analytic approaches for the combination of autonomic and neural activity in the assessment of physiological synchrony*

Ivo V. Stuldreher[1], Joost C.F. de Winter[2], Nattapong Thammasan[3], and Anne-Marie Brouwer[1]

***Abstract*— Physiological synchrony (PS) refers to the similarity in physiological responses of two or more individuals and may be an informative source of information in the field of affective computing. Up to now, PS has been assessed using either autonomic measures or neural measures. While in literature multiple physiological channels have already been combined into one composite index for PS assessment, multimodal PS, i.e., using a combination of autonomic and neural channels in a single composite index ('A-N' multimodal), has remained unexplored. A-N multimodal PS is promising for more robust detection of emotionally or cognitively relevant events, as both autonomic and neural activity are sensitive to these events. The aim of this study is (i) to review analytic approaches that have been used to combine multiple physiological channels into one composite index for PS, within the field of autonomic PS and within the field of neural PS, and (ii) to view them in the light of their potential applicability to A-N multimodal PS.**

A literature search was conducted to find studies assessing PS based on a composite index of multiple autonomic channels or multiple channels in EEG recordings. 4 studies were found that assessed PS based on a composite index using multiple autonomic channels and 12 studies assessed PS based on a composite index using multiple EEG channels.

We found that analytic approaches varied between studies. Some averaged over multiple channels after assessing PS separately per channel ($N = 4$), or averaged over channels before assessing PS ($N = 1$), while others used different linear combinations of channels based on spatio-spectral decomposition ($N = 1$) or correlated component analysis (CCA, $N = 8$). CCA finds linear combinations of channels that are maximally correlated between subjects and has up to now been used to assess neural PS. We argue that this method would be most appropriate for use in multimodal PS assessment.

I. INTRODUCTION

Events that are emotionally or cognitively relevant to an individual induce autonomic responses, such as changes in heart rate or electrodermal activity, as well as responses of the central nervous system (brain signals). These responses are often modest and hard to detect at the level of a single event. Detection is especially hard if one does not want to, or cannot rely on prior knowledge of the time that a relevant event occurred, and if one does not want to collect training data in order to train an individually tailored model for detection of (emotionally) relevant events of the user, as is usually done in the area of affective computing or passive brain-computer interfaces [1]. Within the context of monitoring the mental state of groups of individuals, an approach that could deal with these problems is physiological synchrony (PS). PS refers to the similarity in physiological responses of two or more individuals. It has been studied for autonomic activity and neural activity.

Research on autonomic PS started halfway in the 20th century. Researchers started exploring simultaneous physiological data collection from more than one individual with the aim of studying interpersonal interactions [2]. From then onwards, autonomic PS has been used to study interactions between romantic couples, parent-child dyads and team mates, and has been associated with empathy, relationship quality and team-performance [3], [4], [5].

Neural PS was first monitored in the mid-1960s. Duane and Behrendt [6] recorded neural activity from two participants simultaneously in an attempt to show interpersonal physiological communication. Since then, neural PS has, among others, been shown to be a correlate of classroom engagement [7] and a predictor of expressions of interest and preference during popular television content [8].

Up to now, research on PS has either focused on PS in autonomic activity or on PS in neural activity, while multimodal PS, including both types of signals (from now on referred to as 'A-N' multimodal) might be of added value. We found two studies in which neural activity was monitored in parallel with autonomic PS [9], [10]. However, in these studies, neural activity was only recorded in one of the pair of individuals, as the researchers were interested in neural activity of clients during therapist-client interactions during times of high autonomic PS.

Both autonomic and neural responses can distinguish between affective and neutral stimuli across visual and auditory modalities [11], [12], [13], [14] and autonomic and neural responses can both identify stressful situations [15], [16], [17], [18]. Therefore, multimodal synchrony assessment, based on both neural and autonomic channels, could lead to more robust detection of emotionally or cognitively relevant events.

To enable A-N multimodal PS, an overview of analytic approaches that might be suited for the combination of neural and autonomic measures into a single composite index, for which overall PS can be assessed, is needed. PS literature can provide us with methods enabling the combination of multiple physiological channels into one index for PS assessment, that may be used for this aim. Developments in methodological approaches have led to increasingly ad-

*This work was supported by The Netherlands Organization for Scientific Research (NWA Startimpuls 400.17.602)

[1]Department of Perceptual and Cognitive Systems, TNO, 3769 DE Soesterberg, The Netherlands

[2]Department of Cognitive Robotics, Delft University of Technology, 2600 AA Delft, The Netherlands

[3]Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente, 7500 AE Enschede, The Netherlands

Corresponding author: Anne-Marie Brouwer, anne-marie.brouwer@tno.nl

vanced methods to assess PS. Early studies used zero-order correlations between electrodermal activity and heart rate of therapist and client [19], [20], [2]. Since then, more advanced methods have been used to quantify interpersonal PS, such as wavelet analysis or dynamic systems modeling [21], [22], [23]. The combination of multiple autonomic physiological measures into one composite index was also explored, to provide a single measure of PS. Levenson and Gotmann [4] argued that a composite index provides a better test in characterizing the interpersonal relation in a dyad than separate measures of PS. In neural PS studies, EEG is typically recorded using multiple electrodes. Researchers have used methods to combine these data streams as well.

Existing reviews on analytic approaches for PS [24], [25] do not cover approaches enabling the combination of multiple channels into one composite index, but focus on the assessment of PS between just two channels. The literature review on autonomic PS from Palumbo et al. [26] includes studies using a composite index for PS assessment. However, Palumbo et al. do not focus on the analytic approaches of these studies, do not include approaches from neural literature and do not view the approaches from included studies in the light of potential A-N multimodal applications. The current literature study is aimed at (i) providing an overview of analytic approaches that are used to combine multiple channels into one composite index for PS assessment, within the field of autonomic PS and within the field of neural PS, and (ii) to view them in the light of their potential applicability to A-N multimodal PS.

II. METHODS

To find analytic approaches used to combine multiple *autonomic* measures into a single composite index for PS assessment, the systematic review on autonomic PS from Palumbo et al. [26] was chosen as a start. This review provides an overview of studies monitoring autonomic PS, published in peer-reviewed journals before November 2015. Studies using a composite index for PS assessment (see Appendix F in [26], which summarizes studies according to physiological measures, context, findings and results) were reviewed in the present study. Studies were included if interpersonal PS was assessed using a composite index based on more than one autonomic measure. References of the included studies and citations in the included studies were also reviewed.

For providing an overview of the analytic approaches used to assess *neural* PS, a systematic search was conducted. The goal of this search was to find a comprehensive, but representative sample of studies that monitored multi-channel EEG of multiple participants and assessed their interpersonal PS using a composite index. Using Scopus, a search was conducted with the following search terms.

```
TITLE-ABS-KEY((brain-to-brain OR interbrain
OR inter-brain OR hyperbrain OR intersubject
OR inter-subject) AND (electroencephalogra*
OR eeg) AND (synchron* OR correlation OR
coupling))
```

Based on title, abstract and keywords, studies were only selected if neural PS between multiple human subjects was assessed using a composite index based on multi-channel EEG. The literature study was performed in February 2019.

III. RESULTS

A. Approaches from autonomic literature

From the 61 studies reviewed in [26], only three studies assessed PS based on a composite index. Others only measured one physiological channel ($N = 38$) or measured more than one physiological channel, but assessed PS separately per channel ($N = 20$). After reviewing references in the included studies and citations of the included studies, one more study using a composite index was included, resulting in four studies in total. Table I summarizes the studies according to purpose, sample, physiological measures and the composite index.

The four studies used similar combinations of physiological measures in their composite index. All studies used a measure of cardiovascular activity, three studies used a measure of electrodermal activity and three studies included general somatic activity (body movement). The combination of measures resulted in both sympathetic and parasympathetic measures in all studies.

In two of the included studies [4], [27], the composite index for PS assessment was calculated using a bivariate time-series analysis. All physiological channels were averaged over time within 10 second windows. The z-score of these average values was computed, after which the bivariate time-series analysis, following [28, Chapters 23-25], was performed. For each physiological measure, this analysis provided two chi-square values. These values represented the extent to which the physiological responses from one individual of the couple accounted for the variance in the physiological pattern of their partner, beyond the variance accounted for by the physiological pattern of the partner itself. The z-scores of these values were computed and then averaged over all measures as an overall measure of directional PS in that 10 second time-window. Levenson and Gottman [4] used this to study how PS was associated with marital satisfaction.

Marci [29] combined PS in four physiological channels with the cumulative strength of the overall physiological responses, with the aim to measure audience engagement during television commercials. Unfortunately, it was not described how PS in the four physiological channels was combined into one moment-to-moment time-locked composite index for PS.

Walker et al. [30] investigated to what extent PS between team members predicted team performance in a simulated task, using a multiple regression approach. Team performance was estimated based on task difficulty in the first regression step and based on physiological measures in the second step. The authors described three different approaches. In Analyses 1 & 2, the individual parasympathetic and sympathetic measures were used, respectively, to estimate team performance. In Analyses 3 & 4, the normalized

TABLE I: STUDIES COMBINING MULTIPLE AUTONOMIC CHANNELS INTO ONE COMPOSITE INDEX FOR PS ASSESSMENT.*

Reference	Purpose	Sample	Measures	Composite index
[4]	Determined whether PS could be detected in couples, and whether that was predictive of marital satisfaction.	30 married couples	ACT, IBI, PTT, SCL	Bivariate time-series analysis [28, Chapters 23-25] separately per channel, averaged over all channels.
[27]	Tested whether there was a relationship between PS and perceived empathy.	31 married participants	ACT, FPA, HR, PTT, SCL	Bivariate time-series analysis [28, Chapters 23-25] separately per channel, averaged over all channels.
[29]	Tested whether PS and arousal level predicted viewers' engagement in advertisements.	27 male viewers in two groups	ACT, EDA, HR, RR	PS was combined with physiological intensity, defined as the cumulative strength of the physiological responses, as a measure of viewer engagement.
[30]	Investigated the relationship between PS and joint team performance in a simulated task.	34 two-person teams	LVET, PEP, RSA	Three regression methods were executed to create a measure of team autonomic activity, using individual, correlated and canonically correlated [31] measures of sympathetic and parasympathetic activity respectively.

Measures. ACT = general somatic activity (body movement), EDA = electrodermal activity, HR = heart rate, IBI = inter-beat interval, LVET = left ventricular ejection time, PEP = pre-ejection period, PTT = pulse transmission time, RR = respiration rate, RSA = respiratory sinus arrhythmia, SCL = skin conductance level.
*Parts adapted from [26]

parasympathetic and sympathetic measures were correlated between team members, producing a team parasympathetic score and a team sympathetic score. In Analysis 5, the parasympathetic and sympathetic scores were combined and correlated, using canonical correlation [31].

B. Approaches from EEG literature

Using the search terms described above, 170 studies were found. From these studies, 12 studies that met the inclusion criteria were selected. Table II summarizes them according to purpose, sample, number of channels and the composite index.

The most simple approach that was followed in order to combine multiple measures in one composite index is to average over multiple channels after assessing synchrony separately per channel. In a study on neural PS in students following classes with different teaching styles, Dikker et al. [7] employed a method called total interdependence [32, TI]. Magnitude squared coherence was computed using the Welch method for six one-to-one paired electrodes from two subjects. TI for one pair of subjects was then obtained by averaging over all six electrodes and subject-to-group TI was obtained by averaging over all pairwise combinations of one subject with the other subjects. To validate if TI reflects entrainment to external stimuli, student-to-group TI values were compared between auditory tones and teaching styles with a single-source auditory input (e.g., lecture or reading aloud), hypothesizing that values would be similar. Student-to-group TI was numerically similar in response to tones as in response of the single-source teaching styles.

Lindenberger et al. [33] selected two synchronization measures to study PS in pairs of guitarists: the phase locking index, as a within-brain phase synchronization measure, and interbrain phase coherence (IPC), as a between-brain phase synchronization measure. EEG time series divided into 3 s epochs were transformed into a complex time-frequency signal, for frequencies up to 20 Hz, using a complex Gabor expansion function. IPC values from six fronto-central electrode pairs were averaged in time-frequency domain as measures of PS. Mean IPC values three standard deviations above baseline were considered as statistically significant.

PS of a composite index can also be quantified by averaging over channels before assessing PS. Kinreich et al. [34] computed the Spearman correlation over the time signal of the Stockwell transform frequency spectrum, for each frequency bin, averaged over electrodes within regions of interest (frontal, parietal, temporoparietal, occipital) in two romantic partners and compared this to stranger-dyads. The dyadic correlation values for each frequency bin and region of interest were averaged over two groups (partners and strangers).

Other linear approaches have also been explored. Zamm et al. [35] reduced the dimensionality of the multi-channel EEG recordings to a single dimension using spatio-spectral decomposition [36, SSD], which is a linear spatial decomposition filter. SSD finds a linear filter that maximizes the variance of the signal at peak frequency, while minimizing the variance of the noise at the neighboring frequency bins. After spatial filtering, correlations between the amplitude envelopes of two pianists were calculated as a measure of PS. The observed correlations were compared with a chance distribution of white-noise correlations and observed amplitude envelope correlations were found to be higher than the 95% chance estimates.

Eight studies [37], [8], [38], [39], [40], [41], [42], [43] evaluated intersubject correlation based on correlated component analysis [37, CCA]. The method was designed to find linear combinations of channels that are maximally correlated in time, to identify distributed sources of neural activity. Formally, the approach seeks to maximize the Pear-

TABLE II: STUDIES COMBINING MULTIPLE EEG CHANNELS INTO ONE COMPOSITE INDEX FOR PS ASSESSMENT.

Reference	Purpose	Sample	No. channels*	Composite index
[39]	Determined the conditions under which multisensory stimulation would benefit or hinder the retrieval of everyday experiences, with PS as a marker of enhanced stimulus processing.	88 participants	64	CCA [37], linear combination of channels resulting in maximum ISC
[42]	Investigated whether similarity of EEG responses across subjects to educational videos would be a sensitive measure of knowledge acquisition.	39 participants	64	CCA [37], linear combination of channels resulting in maximum ISC
[7]	Identified whether PS could be a neural markers of group engagement during dynamic real-world group interactions.	12 students	6	Averaged over all channels after calculating PS separately per channel using TI [32]
[37]	Identified brain areas marked by high levels of correlation within and between subjects, evoked by short film clips.	20 participants	64	CCA, linear combination of channels resulting in maximum ISC
[8]	Investigated whether PS was predictive of expressions of interest and viewership during a popular television series.	16 participants	64	CCA [37], linear combination of channels resulting in maximum ISC
[38]	Explored whether PS would predict attentional engagement to a naturalistic narrative stimulus.	76 participants	64	CCA [37], linear combination of channels resulting in maximum ISC
[34]	Compared PS during a male-female naturalistic social interaction between romantic couples and strangers.	24 romantic couples & 25 stranger dyads	32	SC between SP was averaged over channels, after which dyadic correlation values were computed.
[33]	Investigated PS in pairs of guitarists playing a short melody together.	8 pairs of guitarists	16	PLI and IPC values from six fronto-central electrode pairs were averaged in time-frequency domain.
[40]	Investigated whether PS as a measure of auditory attention could distinguish between patients and healthy controls.	20 participants with disorders of consciousness & 14 controls	37	CCA [37], linear combination of channels resulting in maximum ISC
[43]	Investigated whether PS to naturalistic video stimuli decreases with maturity as a marker of neural development.	114 participants (main) & 202 participants (replication)	105	CCA [37], linear combination of channels resulting in maximum ISC
[41]	Determined whether student PS can be quantified in a real-time manner based on portable EEG recordings in a classroom.	28 participants in 4 groups	14	CCA [37], linear combination of channels resulting in maximum ISC
[35]	Explored PS between two pianists performing a musical duet.	1 pair of pianists	24	Channels were combined using SSD [36], after which amplitude envelope correlations were calculated.

Methods. CCA = correlated component analysis, IPC = interbrain phase coherence, PLV = phase locking value measure, SC between SP = Spearman correlation between spectral powers, SSD = spatio-spectral decomposition, TI = total interdependence.
*Represents the No. of channels used in the composite index for synchrony assessment, not the total No. of monitored channels.

son Product Moment Correlation Coefficient. CCA is similar to the more familiar principal component analysis, as both methods project data on a common subspace, except that projections of CCA capture maximal correlation between datasets instead of maximal variance within a dataset. To obtain a measure of subject-to-group PS, data from a single subject was projected on the component vectors. Then, PS of the group was calculated as the correlation coefficient of these projections separately for each component and averaged over all possible subject-pairs involving a single subject. The result is a time-locked, moment-to-moment measure of PS. To check significance of the correlated components, Dmochowski et al. [37] employed a permutation test approach [44]. Correlations were computed with a data set of which 5 s blocks were randomly shuffled in time. The three component correlations were significantly larger than chance levels for 33%, 23% and 10% of movie time for a short, arousing film clip. Similar, Cohen and Parra [39] compared PS of the correlated components to PS in 100 phase-randomized surrogate data-sets [45]. The first three components were selected, as correlations in the weaker components were not always significantly different from chance levels.

IV. DISCUSSION

This literature study was aimed at (i) providing an overview of analytic approaches that are used to combine multiple physiological channels into one composite index for PS assessment and (ii) to view them in the light of their potential applicability to A-N multimodal PS, which will be done below.

Both for autonomic channels and neural channels, PS is most often assessed separately per channel, if multiple channels are assessed at all. The analytic approaches that

are used to combine multiple channels in a composite index vary in complexity and potential suitability for A-N multimodal applications. The simplest indices average over multiple physiological channels after assessing PS separately per channel, such as [7], [4], [27], [33], or average over multiple physiological channels before assessing PS, such as [34]. Averaging over multiple channels can be advantageous, for example to assess synchrony based on multiple electrode readings within a region of interest of the brain or to obtain an overall measure of brain-to-brain synchrony. However, when stimulus-response latency varies strongly between variables, averaging over multiple physiological channels is not appropriate. This is the case when neural and autonomic measures would be considered. For instance, neural event-related potential latencies are in the order of tens or at most a few hundreds of ms, while response latency exceeds 1000 ms for skin conductance responses.

Other linear combinations of channels are not all suited for A-N multimodal applications either. Zamm et al. [35] reduced multi-channel EEG to a single dimension using SSD, which maximizes variance at peak frequency, while reducing variance in neighboring frequency bins. As the frequency power spectrum varies greatly between autonomic and neural measures, SSD is not appropriate for A-N multimodal applications.

The CCA [37], on the other hand, seems appropriate for the exploration of A-N multimodal PS. Data is projected on a subset of data for which intersubject PS is maximized. The method itself thus selects the appropriate combination of channels to maximize PS. Note that for a small number of participants, CCA could lead to misleading results, as it could provide a subset of data that maximizes PS only within the small population. However, the observation that neural PS in a small sample of participants predicted expressions of interest of the larger audience with higher accuracy than the expressions of interest of the small sample [8], suggests that CCA can provide valid results even for small samples of participants. Our next step is to exploit and test this method to analyze an A-N multimodal dataset we recently collected from participants who were presented with the same audio but differed in attentional focus (see text box).

We end by noting two limitations of the current literature review. Firstly, we did not do a targeted search for combining multiple autonomic and/or neural measures in intra-person synchrony. Such a search may still provide new insights as to methodology to reach A-N multimodal composite indices for PS assessment. Secondly, for neural measures, the current literature review only examined analytic approaches used to assess multi-channel electroencephalographic PS. Analytic approaches used for assessing PS monitored with other neuroimaging techniques, such as fMRI or MEG, might also be suited for A-N multimodal applications.

EEG, ECG and electrodermal activity signals were obtained from participants who either attended to an audiobook or to interspersed auditory events which were emotional sounds from the IADS-2 [46], high and low pitched beeps that attending participants needed to keep track of [47] and, at the end, the instruction to sing a song aloud after the subsequent auditory countdown reached 0 [48]. We expect that all of these types of events will generate attentional and emotional (arousal) processes, especially in the participants who attend to them, and that multimodal PS will be most robust and outperform unimodal PS in detecting the occurrence of these events and distinguish between the two groups of participants. While we expect similarities across responses to the different types of relevant events, we also expect some effect of the type of event on the strength of PS between different measures. The cognitive event (interval with high and low beeps) is expected to especially elicit workload related responses, which are especially clear in EEG [49] such that EEG synchrony may be stronger than autonomic synchrony. Emotional events (the task to sing a song and the emotional sounds) may especially generate autonomic synchrony. If the measures showing PS depend very strongly on the type of event, it may not be advantageous to have one single multimodal composite index after all, but to have different composite indices simultaneously that might each be sensitive for a certain type of relevant event. With our data, we will be able to decide whichever situation is the case.

REFERENCES

[1] A.-M. Brouwer, T. O. Zander, J. B. Van Erp, J. E. Korteling, and A. W. Bronkhorst, "Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls," *Frontiers in Neuroscience*, vol. 9, p. 136, 2015.

[2] A. DiMascio, R. W. Boyd, M. Greenblatt, and H. C. Solomon, "The psychiatric interview: a sociophysiologic study." *Diseases of the Nervous System*, vol. 16, no. 1, pp. 4–9, 1955.

[3] C. D. Marci, J. Ham, E. Moran, and S. P. Orr, "Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy," *The Journal of Nervous and Mental Disease*, vol. 195, no. 2, pp. 103–111, 2007.

[4] R. W. Levenson and J. M. Gottman, "Marital interaction: physiological linkage and affective exchange." *Journal of Personality and Social Psychology*, vol. 45, no. 3, p. 587, 1983.

[5] A. N. Elkins, E. R. Muth, A. W. Hoover, A. D. Walker, T. L. Carpenter, and F. S. Switzer, "Physiological compliance and team performance," *Applied Ergonomics*, vol. 40, no. 6, pp. 997–1003, 2009.

[6] T. D. Duane and T. Behrendt, "Extrasensory electroencephalographic induction between identical twins." *Science*, 1965.

[7] S. Dikker, L. Wan, I. Davidesco, L. Kaggen, M. Oostrik, J. McClintock, J. Rowland, G. Michalareas, J. J. Van Bavel, M. Ding *et al.*, "Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom," *Current Biology*, vol. 27, no. 9, pp. 1375–1380, 2017.

[8] J. P. Dmochowski, M. A. Bezdek, B. P. Abelson, J. S. Johnson, E. H. Schumacher, and L. C. Parra, "Audience preferences are predicted by temporal reliability of neural processing," *Nature Communications*, vol. 5, p. 4567, 2014.

[9] T. Stratford, S. Lal, A. Meara *et al.*, "Neurophysiology of therapeutic alliance," *Gestalt Journal of Australia and New Zealand*, vol. 5, no. 2, p. 19, 2009.

[10] T. Stratford, S. Lal, and A. Meara, "Neuroanalysis of therapeutic alliance in the symptomatically anxious: the physiological connection revealed between therapist and client," *American Journal of Psy-*

chotherapy, vol. 66, no. 1, pp. 1–21, 2012.

[11] M. M. Bradley and P. J. Lang, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 2, pp. 204–215, 2000.

[12] B. N. Cuthbert, H. T. Schupp, M. M. Bradley, N. Birbaumer, and P. J. Lang, "Brain potentials in affective picture processing: covariation with autonomic arousal and affective report," *Biological Psychology*, vol. 52, no. 2, pp. 95–111, 2000.

[13] D. T. Hettich, E. Bolinger, T. Matuz, N. Birbaumer, W. Rosenstiel, and M. Spüler, "Eeg responses to auditory stimuli for automatic affect recognition," *Frontiers in Neuroscience*, vol. 10, p. 244, 2016.

[14] P. Lang and M. M. Bradley, "The international affective picture system (iaps) in the study of emotion and attention," *Handbook of Emotion Elicitation and Assessment*, vol. 29, 2007.

[15] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and galvanic skin response signals," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 2013, pp. 209–214.

[16] J. Choi and R. Gutierrez-Osuna, "Using heart rate monitors to detect mental stress," in *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*. IEEE, 2009, pp. 219–223.

[17] G. Jun and K. G. Smitha, "Eeg based stress level identification," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016, pp. 003 270–003 274.

[18] M. Haak, S. Bos, S. Panic, and L. Rothkrantz, "Detecting stress using eye blinks and brain activity from eeg signals," *Proceeding of the 1st Driver Car Interaction and Interface (DCII 2008)*, pp. 35–60, 2009.

[19] R. Coleman, M. Greenblatt, and H. C. Solomon, "Physiological evidence of rapport during psychotherapeutic interviews." *Diseases of the Nervous System*, vol. 17, no. 3, p. 71, 1956.

[20] A. DiMascio, R. W. Boyd, and M. Greenblatt, "Physiological correlates of tension and antagonism during psychotherapy: A study of" interpersonal physiology"," *Psychosomatic Medicine*, vol. 19, no. 2, pp. 99–104, 1957.

[21] G. Quer, J. Daftari, and R. R. Rao, "Heart rate wavelet coherence analysis to investigate group entrainment," *Pervasive and Mobile Computing*, vol. 28, pp. 21–34, 2016.

[22] J. L. Helm, D. Sbarra, and E. Ferrer, "Assessing cross-partner associations in physiological responses via coupled oscillator models." *Emotion*, vol. 12, no. 4, p. 748, 2012.

[23] E. Ferrer and J. L. Helm, "Dynamical systems modeling of physiological coregulation in dyadic interactions," *International Journal of Psychophysiology*, vol. 88, no. 3, pp. 296–308, 2013.

[24] K. M. Gates and S. Liu, "Methods for quantifying patterns of dynamic interactions in dyads," *Assessment*, vol. 23, no. 4, pp. 459–471, 2016.

[25] J. L. Helm, J. G. Miller, S. Kahle, N. R. Troxel, and P. D. Hastings, "On measuring and modeling physiological synchrony in dyads," *Multivariate Behavioral Research*, vol. 53, no. 4, pp. 521–543, 2018.

[26] R. V. Palumbo, M. E. Marraccini, L. L. Weyandt, O. Wilder-Smith, H. A. McGee, S. Liu, and M. S. Goodwin, "Interpersonal autonomic physiology: A systematic review of the literature," *Personality and Social Psychology Review*, vol. 21, no. 2, pp. 99–141, 2017.

[27] R. W. Levenson and A. M. Ruef, "Empathy: A physiological substrate." *Journal of Personality and Social Psychology*, vol. 63, no. 2, p. 234, 1992.

[28] J. M. Gottman, *Time-series analysisa comprehensive introduction for social scientists*. Cambridge University Press, 1981, no. 519.55 G6.

[29] C. D. Marci, "A biologically based measure of emotional engagement: Context matters," *Journal of Advertising Research*, vol. 46, no. 4, pp. 381–387, 2006.

[30] A. D. Walker, E. R. Muth, F. S. Switzer III, and P. J. Rosopa, "Predicting team performance in a dynamic environment: A team psychophysiological approach to measuring cognitive readiness," *Journal of Cognitive Engineering and Decision Making*, vol. 7, no. 1, pp. 69–82, 2013.

[31] B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics (4th ed.)*. Allyn & Bacon, 2001.

[32] X. Wen, J. Mo, and M. Ding, "Exploring resting-state functional connectivity with total interdependence," *NeuroImage*, vol. 60, no. 2, pp. 1587–1595, 2012.

[33] U. Lindenberger, S.-C. Li, W. Gruber, and V. Müller, "Brains swinging in concert: cortical phase synchronization while playing guitar," *BMC Neuroscience*, vol. 10, no. 1, p. 22, 2009.

[34] S. Kinreich, A. Djalovski, L. Kraus, Y. Louzoun, and R. Feldman, "Brain-to-brain synchrony during naturalistic social interactions," *Scientific Reports*, vol. 7, no. 1, p. 17060, 2017.

[35] A. Zamm, S. Debener, A.-K. R. Bauer, M. G. Bleichner, A. P. Demos, and C. Palmer, "Amplitude envelope correlations measure synchronous cortical oscillations in performing musicians," *Annals of the New York Academy of Sciences*, 2018.

[36] V. V. Nikulin, G. Nolte, and G. Curio, "A novel method for reliable and fast extraction of neuronal eeg/meg oscillations on the basis of spatio-spectral decomposition," *NeuroImage*, vol. 55, no. 4, pp. 1528–1535, 2011.

[37] J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra, "Correlated components of ongoing eeg point to emotionally laden attention–a possible marker of engagement?" *Frontiers in Human Neuroscience*, vol. 6, p. 112, 2012.

[38] J. J. Ki, S. P. Kelly, and L. C. Parra, "Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli," *Journal of Neuroscience*, vol. 36, no. 10, pp. 3092–3101, 2016.

[39] S. S. Cohen and L. C. Parra, "Memorable audiovisual narratives synchronize sensory and supramodal neural responses," *eNeuro*, pp. ENEURO–0203, 2016.

[40] I. Iotzov, B. C. Fidali, A. Petroni, M. M. Conte, N. D. Schiff, and L. C. Parra, "Divergent neural responses to narrative speech in disorders of consciousness," *Annals of Clinical and Translational Neurology*, vol. 4, no. 11, pp. 784–792, 2017.

[41] A. T. Poulsen, S. Kamronn, J. Dmochowski, L. C. Parra, and L. K. Hansen, "Eeg in the classroom: Synchronised neural recordings during video presentation," *Scientific Reports*, vol. 7, p. 43916, 2017.

[42] S. S. Cohen, J. Madsen, G. Touchan, D. Robles, S. F. Lima, S. Henin, and L. C. Parra, "Neural engagement with online educational videos predicts learning performance for individual students," *bioRxiv*, p. 253260, 2018.

[43] A. Petroni, S. S. Cohen, L. Ai, N. Langer, S. Henin, T. Vanderwal, M. P. Milham, and L. C. Parra, "The variability of neural responses to naturalistic videos change with age and sex," *eNeuro*, vol. 5, no. 1, pp. ENEURO–0244, 2018.

[44] R. A. Fisher, *The design of experiments*. Oliver And Boyd; Edinburgh; London, 1937.

[45] D. Prichard and J. Theiler, "Generating surrogate data for time series with several simultaneously measured variables," *Physical Review Letters*, vol. 73, no. 7, p. 951, 1994.

[46] M. M. Bradley and P. J. Lang, "The international affective digitized sounds (; iads-2): Affective ratings of sounds and instruction manual," *University of Florida, Gainesville, FL, Tech. Rep. B-3*, 2007.

[47] A. L. De Dieuleveult, A.-M. Brouwer, P. C. Siemonsma, J. B. Van Erp, and E. Brenner, "Aging and sensitivity to illusory target motion with or without secondary tasks," *Multisensory Research*, vol. 31, no. 3-4, pp. 227–249, 2018.

[48] A.-M. Brouwer and M. A. Hogervorst, "A new paradigm to induce mental stress: the sing-a-song stress test (ssst)," *Frontiers in Neuroscience*, vol. 8, p. 224, 2014.

[49] M. A. Hogervorst, A.-M. Brouwer, and J. B. Van Erp, "Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload," *Frontiers in Neuroscience*, vol. 8, p. 322, 2014.

C

Stimuli

C.1. Affective sounds (IADS-2)

The IADS-2 database consists of 165 sounds with normative ratings of valence, arousal and dominance, obtained using the Self-Assessment Manikin (SAM, Lang, 1980) with at least 100 participants per sound. For this study, three sets of 12 sounds each were selected: neutral (mean valence, low arousal), pleasant (high valence, high arousal) and unpleasant (low valence, high arousal). Sounds were selected not only on the mean rating across participants, it was also aimed to select sounds with low variation in ratings between participants (i.e., low standard deviation). Tables C.1, C.2 and C.3 show the mean and standard deviation of normative ratings of valence, arousal and dominance for neutral, pleasant and unpleasant sounds, respectively.

Table C.1 Neutral IADS-2 sounds, summarized according to their number, name and normative ratings of valence, arousal and dominance.

No.	Name	Valence		Arousal		Dominance	
		Mean	SD	Mean	SD	Mean	SD
246	HeartBeat	4.89	1.69	4.48	2.58	5.35	2.07
262	Yawn	5.2	1.31	3.12	1.92	4.92	1.87
373	Paint	5.12	1.54	5.25	1.74	5.57	1.76
376	Lawnmower	5.69	1.32	4.23	1.69	5.86	1.55
382	Shovel	4.55	1.25	4.51	1.73	4.87	1.6
627	Rain1	4.73	1.67	4.61	1.96	4.59	1.87
698	Rain2	5.31	1.97	4.46	2.53	5.03	2.16
700	Toilet	5.12	1.67	3.96	2.14	5.91	1.98
708	Clock	4.46	1.68	3.22	2.25	4.86	2.33
720	BrushTeeth	4.91	1.54	4.13	2.01	5.61	2.14
723	Radio	4.41	1.53	4.22	1.93	5.22	2.02
728	Paper1	4.86	1.34	4.17	1.93	5.58	1.71

Table C.2 Pleasant IADS-2 sounds, summarized according to their number, name and normative ratings of valence, arousal and dominance.

No.	Name	Valence		Arousal		Dominance	
		Mean	SD	Mean	SD	Mean	SD
110	Baby	7.64	2.1	6.03	2.1	6.14	1.88
200	EroticCouple	6.71	1.78	7.49	1.52	6.29	2.03
201	EroticFem1	7.87	1.62	8.13	1.36	7.48	1.8
202	EroticFem2	7.9	1.48	7.67	1.8	7.3	1.85
205	EroticFem3	6.84	1.98	6.96	2.04	6.41	1.84
311	Crowd2	7.65	1.58	7.12	1.58	6.09	2.18
352	SportsCrowd	7.17	1.97	7.07	1.97	5.77	2.08
353	Baseball	7.38	1.53	6.62	1.63	6.04	1.88
360	Rollercoaster	6.94	2.25	7.54	2.25	4.73	2.39
365	Party	6.97	1.9	6.32	1.9	5.73	1.76
366	Casino1	7.09	1.73	6.26	1.73	6.08	2.19
367	Casino2	7.33	1.74	6.72	1.74	6.41	1.98
415	Countdown	6.46	1.67	6.55	1.67	4.8	2.25
717	SlotMachine2	7.32	1.64	6.56	1.64	6.39	2.3

Table C.3 Unpleasant IADS-2 sounds, summarized according to their number, name and normative ratings of valence, arousal and dominance.

No.	Name	Valence		Arousal		Dominance	
		Mean	SD	Mean	SD	Mean	SD
115	Bees	1.98	1.32	7.2	1.91	2.58	1.75
255	Vomit	1.71	1.76	7.29	1.91	2.64	1.58
260	BabiesCry	2.28	1.54	6.44	2.09	3.95	2.37
276	FemScream2	2.04	1.38	7.65	1.65	3.16	2.15
277	FemScream3	1.85	1.41	7.55	1.93	2.72	2.04
278	ChildAbuse	1.78	1.57	7.16	2.11	3.96	2.49
279	Attack1	1.9	1.23	7.9	1.31	2.84	2.18
284	Attack3	2.51	1.85	6.63	2.14	3.71	2.16
285	Attack2	1.87	1.5	7.49	1.91	2.68	1.92
286	Victim	2.04	1.46	7.69	1.69	2.96	2.33
290	Fight1	1.92	1.54	7.38	1.76	3.7	2.15
292	MaleScream	2.6	1.74	6.97	1.7	3.49	1.99
422	TireSkids	2.72	1.6	7.01	1.77	2.84	1.84
424	CarWreck	2.74	1.94	7.68	1.74	2.82	2.05
600	BikeWreck	2.73	1.8	6.96	1.89	3.43	2.02

C.2. Working memory task (WMT)

An auditory working memory task (WMT) as presented in (De Dieuleveult et al., 2018) was adjusted to fit this study's needs. In the current study, participants were presented with low and high tones (250 Hz for 100 ms; 1 kHz for 100 ms), one of which was presented every two seconds for a trial duration of 30 seconds. Sequences of tones were randomly produced, by selecting one of the two tones at each instance with a probability of 50%. In Table C.4 the 27 sequences of 15 tones as presented in this study are shown.

Table C.4 Specification of working memory task (WMT) trials 001 to 027. In each trial, either a low (250 Hz, 100 ms) or high (1 kHz, 100 ms) tone was presented every two seconds for a total trial duration of 30 seconds (thus resulting in 15 tones per trial).

No.	Tone 01 - 15														
001	high	low	low	high	high	high	low	high	high	low	high	low	low	high	high
002	high	high	low	high	high	high	high	high	low	high	low	high	low	low	low
003	low	high	high	low	high	low	low	low	high	high	low	low	low	high	high
004	high	low	high	high	low	low	low	high	low	high	low	high	low	high	high
005	high	high	high	low	low	low	high	low	high	low	high	low	low	low	high
006	low	low	high	high	high	high	low	high	high	low	high	low	low	high	high
007	high	low	high	low	low	low	low	high	low	high	low	high	low	high	high
008	high	low	low	low	high	low	high	high	high	low	low	low	high	low	high
009	high	high	low	low	low	high	low	high	low	low	low	low	high	high	high
010	low	high	high	low	high	low	low	low	low	low	low	low	low	high	high
011	low	low	low	high	low	low	high	low	low	low	low	low	high	high	high
012	low	low	low	high	low	low	low	high	high	high	low	high	low	high	low
013	high	low	low	high	high	low	high	high	low	low	low	low	high	high	high
014	high	high	low	high	high	low	high	high	high	high	high	low	low	low	low
015	high	low	low	low	low	low	low	high	low	low	high	high	low	low	low
016	low	high	low	high	high	low	low	low	low	low	high	low	low	high	low
017	high	high	low	high	low	low	high	high	high	low	low	high	high	low	low
018	high	low	high	high	high	low	low	low	high	low	high	low	high	low	high
019	high	high	high	low	high	low	low	high	high	low	high	high	high	high	high
020	high	low	low	high	low	low	low	high	high	high	low	low	high	low	low
021	high	low	high	high	high	low	high	high	high	high	high	low	low	high	low
022	low	low	low	high	high	low	low	high	low	high	high	low	low	low	high
023	low	low	low	low	low	low	high	high	low	high	low	low	high	high	low
024	high	low	high	high	high	high	high	low	low	high	low	low	high	high	high
025	low	low	low	high	low	high	high	low	low	low	low	low	high	low	low
026	high	low	low	high	low	high	high	high	low	high	low	high	high	high	low
027	high	low	low	low	low	low	low	low	high	low	high	high	high	low	low

C.3. Interspersed stimulus order and timing

To ensure identical perceptual input for all participating individuals, the stimulus presentation order and timing were identical for all participants. Table C.5 presents the time of stimulus onset and stimulus duration for all the interspersed stimuli.

Table C.5 Stimulus presentation order and timing for the auditory stimuli as presented to the participants. Besides silent baseline (Silence) and white noise baseline (White noise), participants were presented with one narrative stimulus (Zure Koekjes), interspersed with three types of stimuli. Stimuli No. 0 correspond to working memory task (WMT) trials, No. 1 correspond to affective sounds (IADS-2) and No. 2 corresponds to the sing-a-song stress test (SSST).

No.	Name	Time (s)	Duration (s)
3	Silence	0	60
4	White noise	60	60
	Zure Koekjes	120	
1	ChildAbuse	140	6
1	EroticFem2	192	6
1	HeartBeat	246	6
1	Casino2	292	6
0	001	345	30
0	022	428	30
1	Lawnmower	512	6
0	016	572.1	30
1	Attack1	638.1	6
1	Paint	698.1	6
1	BabiesCry	749.1	6
1	Clock	791.1	6
0	007	835.1	30
1	Vomit	906.1	6
1	Baby	950.1	6
0	019	1006.2	30
1	Crowd2	1089.2	6
0	012	1147.2	30
0	010	1226.2	30
1	Shovel	1307.2	6
0	027	1350.2	30
1	EroticFem1	1425.2	6
1	Countdown	1467.2	6
1	FemScream2	1525.2	6
0	003	1573.2	30
0	013	1655.2	30
1	Toilet	1735.2	6
0	023	1783.2	30
0	011	1852.2	30
1	MaleScream	1924.2	6
1	Victim	1974.2	6
1	FemScream3	2026.2	6
1	Rain2	2075.2	6
0	024	2131.2	30
0	020	2196.2	30
0	009	2268.2	30
0	026	2336.2	30
1	Baseball	2408.2	6
1	BrushTeeth	2455.3	6
0	008	2496.3	3
0	002	2575.3	30
0	021	2654.3	30
1	SlotMachine2	2736.3	6
1	Bees	2786.3	6
1	Casino1	2840.3	6
0	015	2894.3	30
1	Radio	2966.3	6
0	025	3013.3	30
1	Rain1	3092.3	6
1	Attack2	3133.4	6
1	TireSkids	3193.4	6
1	Paper1	3247.4	6
0	006	3290.4	30
1	Yawn	3374.4	6
1	Fight1	3424.4	6
0	004	3484.4	30
1	Party	3569.4	6
0	005	3615.4	30
1	SportsCrowd	3691.4	6
1	EroticCouple	3732.4	6
0	018	3776.4	30
0	014	3848.4	30
0	017	3929.4	30
2	Sing-a-Song Test	4000.4	26.7
1	White noise	4087.3	60
1	Silence	4147.3	60

D

Questionnaires

D.1. Self-reported measures of experiment experiences

Self-reported measures of attention, workload and emotion were obtained on a continuous scale from 'not at all' to 'very', mapped to from 0 to 1 in increments of 0.1. Figure D.1 shows the example question that was shown to participants to explain the scale. Below this figure, the statements that were presented to participants are listed.

Figure D.1 Self-reported measures of attention, workload and emotional were provided on a continous scale from 0 to 1, corresponding to 'not at all' (helemaal niet) and 'very' (helemaal wel) respectively.

(Translated from Dutch) The next statements cover your experience during the experiment. For each of the statements presented, click anywhere on the line corresponding to what is most applicable:

- I was distracted by the other audiostream.
- I made mental effort during the experiment.

- I was emotional due to the short sounds, such as screaming, when listening to the audiobook.
- I was stressed by the sing-a-song assignment at the end of the experiment.

(The following statements were presented to narrative-attending participants only)

- I was distracted by the short beeps when listening to the book.
- I was distracted by the short sounds when listening to the book.
- I was distracted by the sing-a-song assignment at the end

(The following statements were presented to short-stimuli-attending participants only)

- I made mental effort when counting tones.
- I was distracted by the audiobook.

D.2. Questions indicative of directed attention

To obtain measures indicative of directed attentional engagement, all participants were asked to answer questions about the interspersed stimuli and narrative of the audiobook. The questions (correct answer in brackets) are shown below.

The next questions cover the short stimuli presented during the audiobook presentation.

- How many tones were on average played in a sequence of the tone-counting assignment? (15)
- How many high tones were on average played in a sequence of the tone-counting assignment? (8)
- How many low tones were on average played in a sequence of the tone-counting assignment? (7)
- How many short sounds, such as screaming, did you hear? (36)
- Describe as many of the short sounds that you heard. (see C.1)

The next questions cover the content of the audiobook.

- At what time was the last call with Mr. van Averdonk's mobile phone? (15:32)
- What is the title on Valerie's business card? (Marketing Manager)
- What address is written on the back of Emma's picture? (Weertjeshof)
- Who was the only person strongly believing in the sour cookies? (Mr. van Keulen)
- What is the estimated revenue of Van Averdonk en Co.? (300 million euros)
- How did Maria meet her husband Willem? (As a nurse in the a hospital where Willem was hospitalized)

- The son of chief inspector Helena Visscher says he has a black eye. Who is guilty according to him? (The neighbor)
- How did Jurgen react on the confession of Maria van Averdonk? (Not in shock)
- What did Helena do to blame Mr. van Keulen for the murder? (She wrote a suicide note on behalf of Mr. van Keulen)
- How old is Emma on the day of the murder of Mr. van Averdonk? (40)

E

Processing physiological responses

E.1. ECG QRS complex detection

ECG signals are inherently nonstationary, meaning that their features are localized in time and frequency. Wavelets can be used to decompose ECG into time-varying frequency components. The raw signal is thus separated into multiple frequency bands, enabling a sparser signal representation better suited for detection of the QRS complex; the most important wave, characterizing the ventricular contractions.

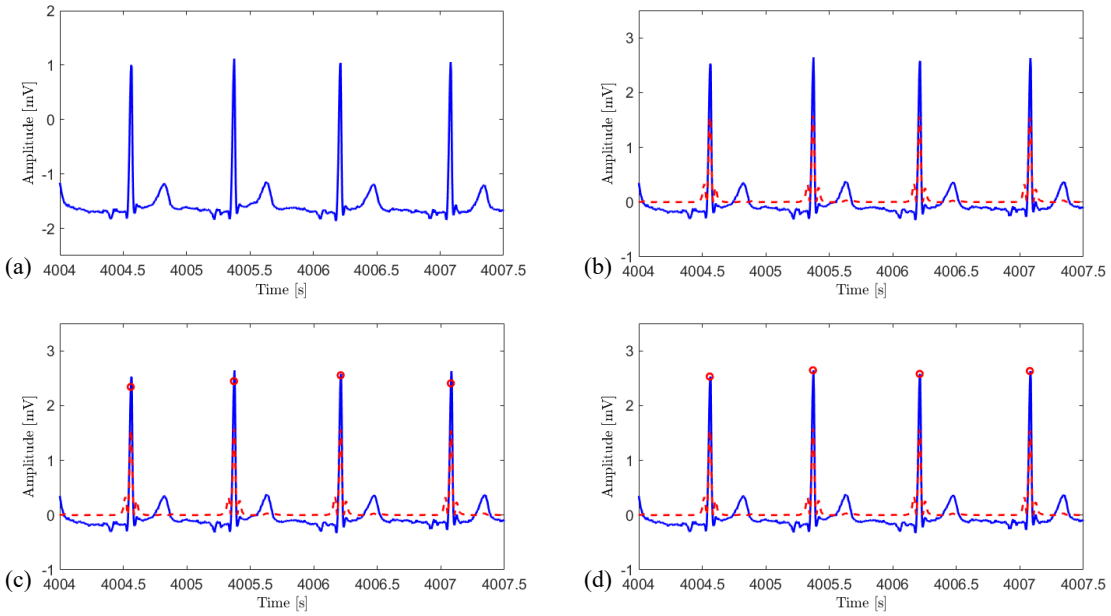


Figure E.1 Processing steps for ECG peak detection on data from a representative participant. (a) Raw ECG signal. (b) Filtered ECG signal (-) and squared version of the frequency localized reconstructed waveform using wavelet coefficients (- -). (c) Uncorrected detection of R peaks (o). (d) Corrected detection of R peaks.

In this study, the maximal overlap discrete wavelet transform (MODWT, Percival and Walden, 2006), an undecimated wavelet transform, was used to enhance R peaks of the ECG signal. The ECG

signal was first decomposed down to level 5 using the default 'sym4' wavelet as implemented in the MATLAB function `modwt`. A frequency localized version of the ECG waveform using wavelet coefficients at scale 4 and 5 was reconstructed using the MATLAB function `imodwt`. The scales correspond to $\left[11.25 \quad 22.5\right]$ Hz and $\left[5.625 \quad 11.25\right]$ Hz respectively and together cover the passband of the ECG signal shown to maximize energy of the QRS complex. R peaks were then detected on the squared version of this signal with the `findpeaks` function. Here, 'minpeakheigth' was set at 0.1 and 'minpeakdistance' was set at 300 ms, as shorter inter-beat intervals - corresponding to a HR higher than 200 bpm - are certainly not realistic in our stationary laboratory setting. As the peak detection algorithm sometimes shows an offset of a few ms compared to the maximum peak value, the peak location was corrected by searching for the maximum value within 10 ms before or after the found peak. Figure E.1 shows these processing steps on ECG data from a representative participant.

E.2. Independent Component Analysis for EEG artifact detection

EEG was processed offline with EEGLAB v14.1.2 for MATLAB (Delorme and Makeig, 2004). For removal of ocular artifacts, independent component analysis (ICA) was conducted (see Matlab script `prepro_eeg_ica`). EEG was first downsampled to 256 Hz and high-pass filtered at 1 Hz, using the standard FIR-filter implement in EEGLAB function `pop_eegfiltnew`. Compared to lower cutoff frequencies, this has shown to work better for ICA (Winkler et al., 2015). Data were also notch-filtered at 50 Hz. Bad channels were removed based on their statistical properties. Channels whose standard deviation exceeded $50\mu V$ were removed. Then, within kept channels, samples exceeding ± 150 μV, as well as samples within 100 ms, were also removed. Bad channels were interpolated using `pop_interp`. Channels were re-referenced to the average channel values. Note that re-referencing is a must for data obtained with the BioSemi ActiveTwo system, as the signals still contain some common-mode noise. Logistic infomax ICA (Bell and Sejnowski, 1995) was performed to localize independent components. The Multiple Artifact Rejection Algorithm (MARA, Winkler et al., 2011) was executed to classify artifactual independent components, i.e., components not reflecting sources of neural activity, but ocular or muscle-related artifacts. MARA is a supervised learning algorithm that learns from expert ratings of 1290 components. It classifies components based on six features from the spatial, spectral and temporal domain. Figure E.2 shows classification of the first 24 independent components from data of a representative participant.

E.3. Matlab code

The Matlab code used for this study is published in the online repository linked below. This repository also contains information how to use the scripts:
`https://gitlab.com/ivostuldreher/matlab-code-thesis`

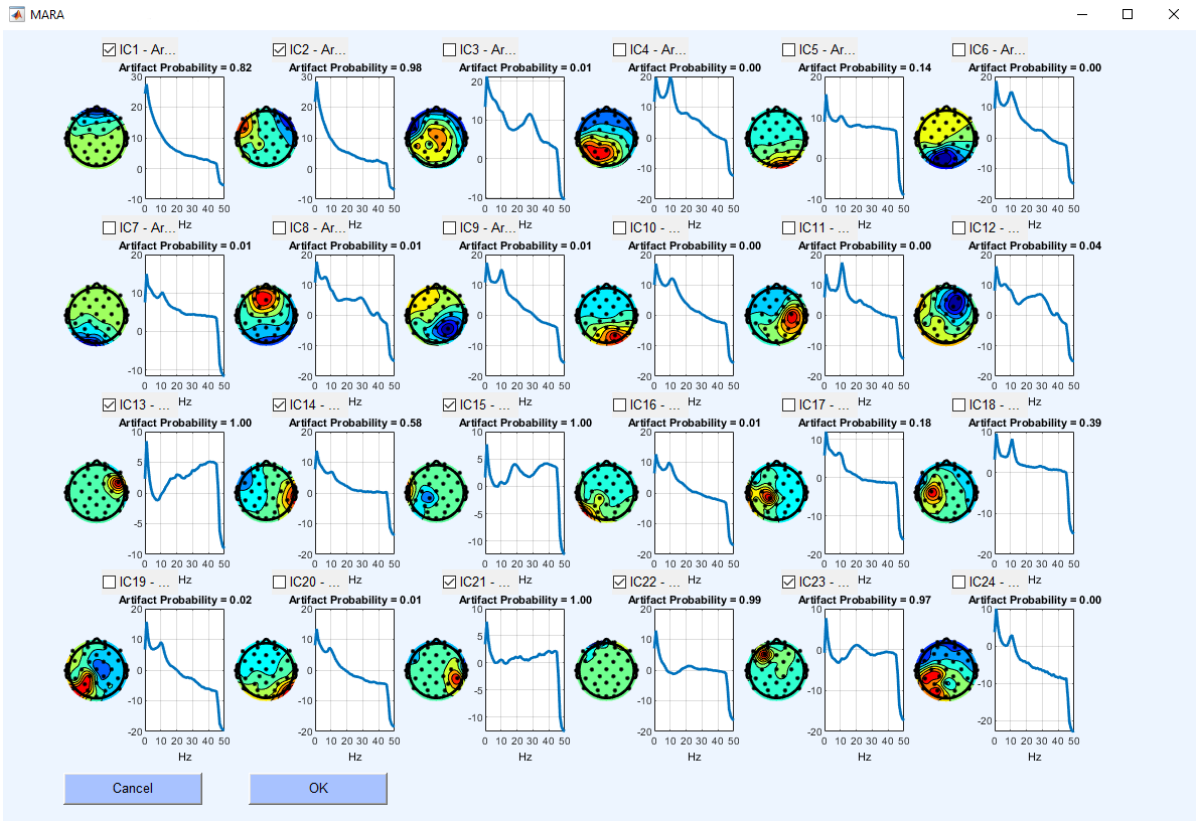


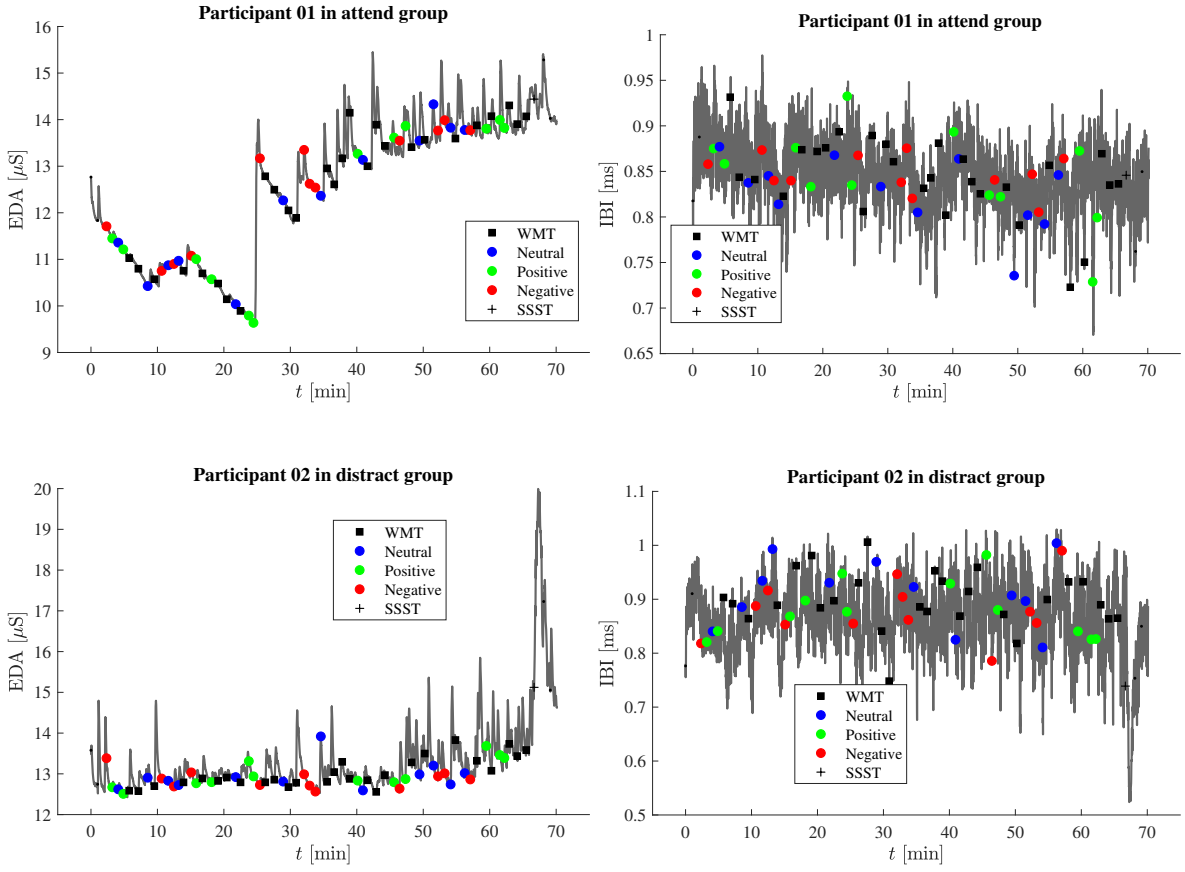
Figure E.2 Illustration of MARA on the first 24 independent components of a representative participants. Marked components 1 and 2 likely correspond to eye-related artifacts, whereas component 13 likely corresponds to a muscle artifact.

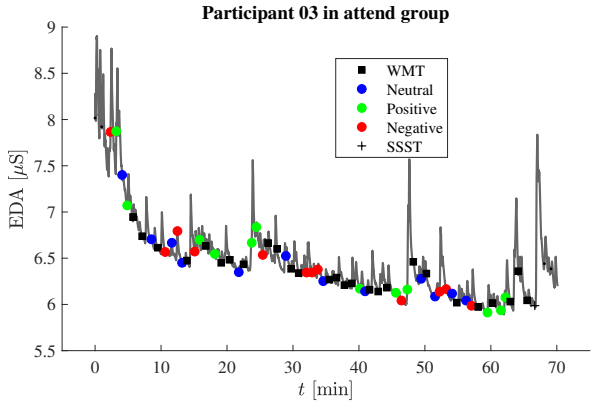
F

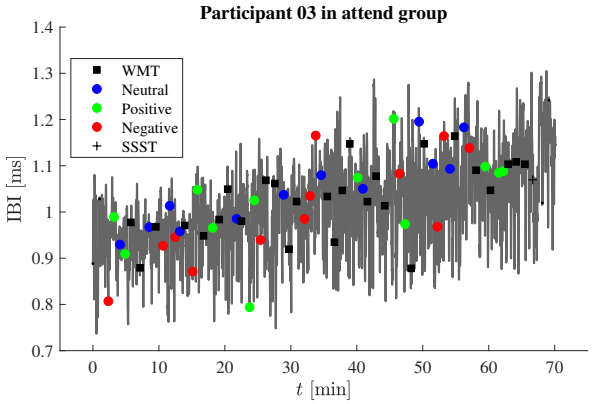
Supplementary results

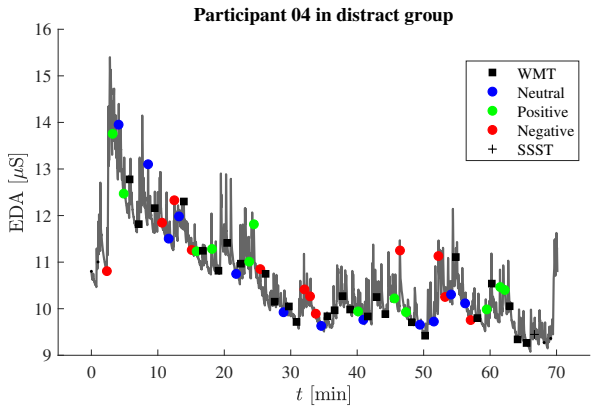
F.1. Individual physiological responses

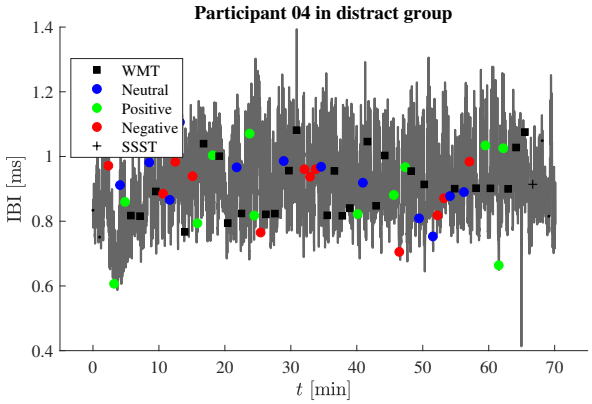
Figure F.1, presented below this section, shows autonomic stimulus response traces - i.e., raw electrodermal activity and inter-beat interval - for each participant across the entire experiment. Markers show onset of interspersed stimuli.

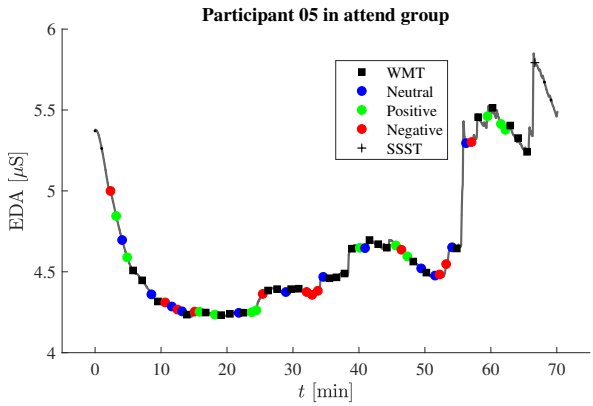


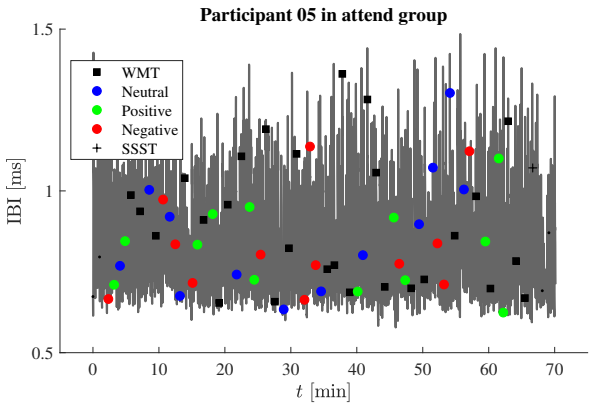

Participant 03 in attend group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

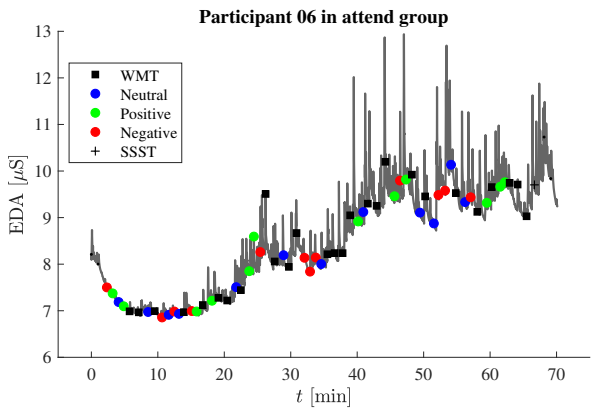

Participant 03 in attend group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

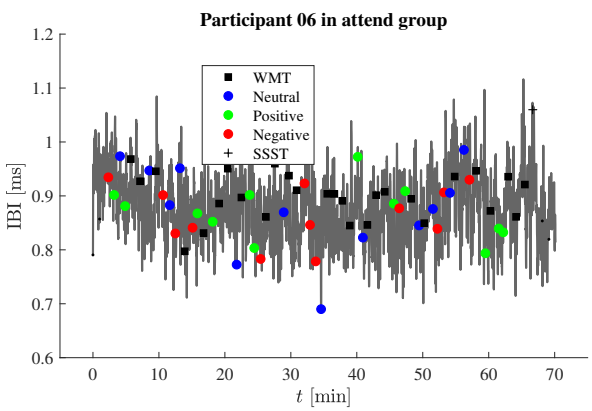

Participant 04 in distract group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

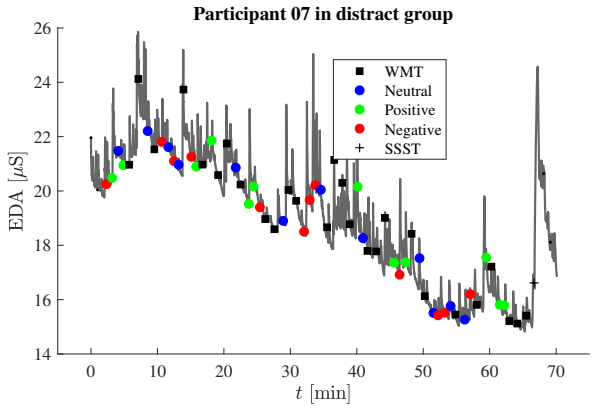

Participant 04 in distract group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

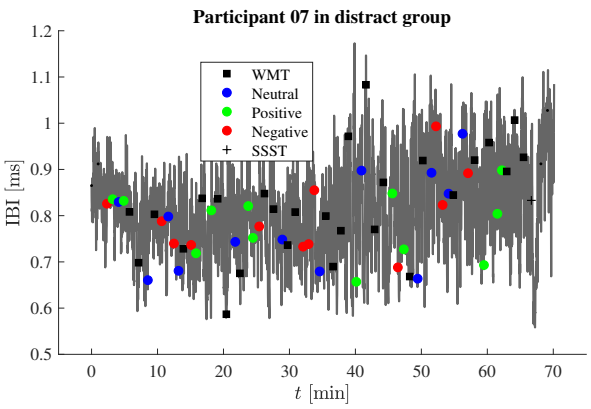

Participant 05 in attend group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

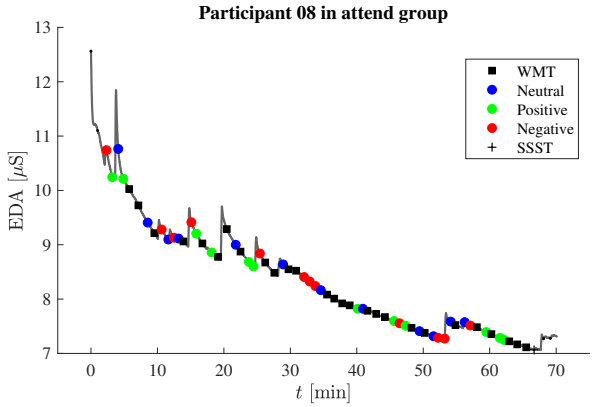

Participant 05 in attend group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

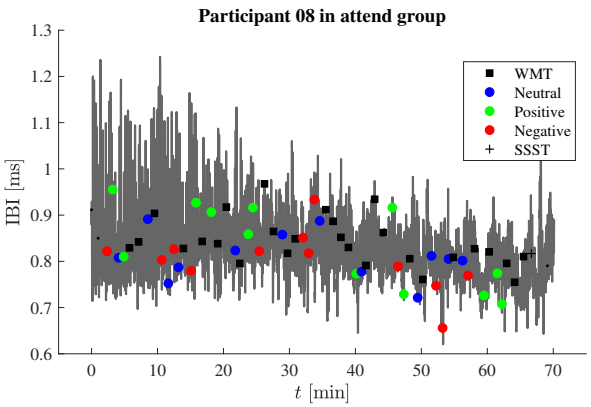

Participant 06 in attend group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

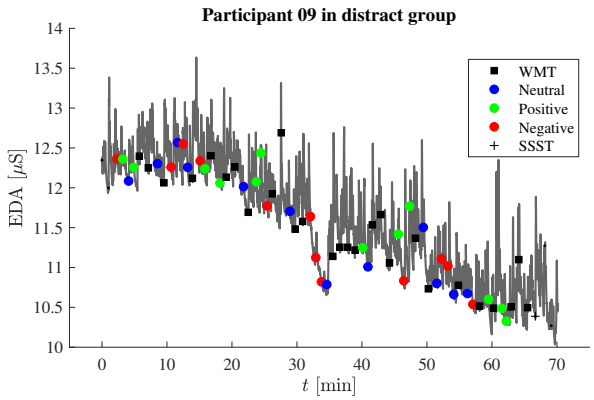

Participant 06 in attend group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

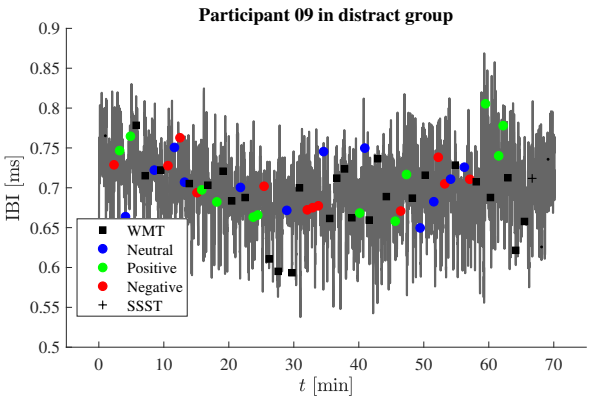

Participant 07 in distract group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

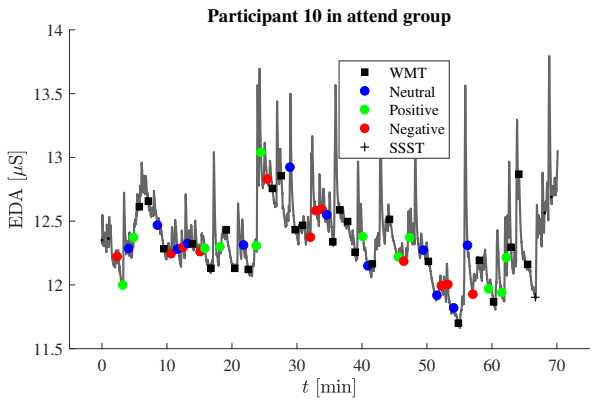

Participant 07 in distract group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

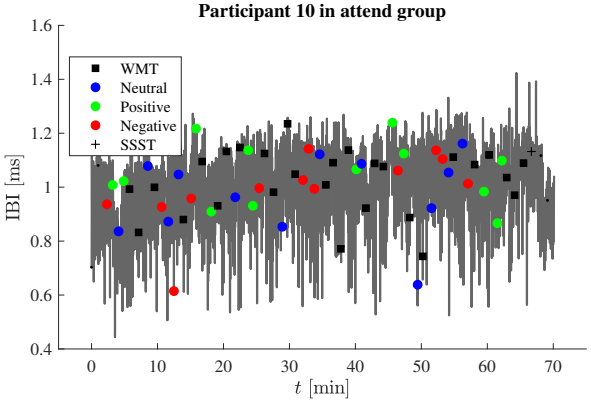

Participant 08 in attend group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

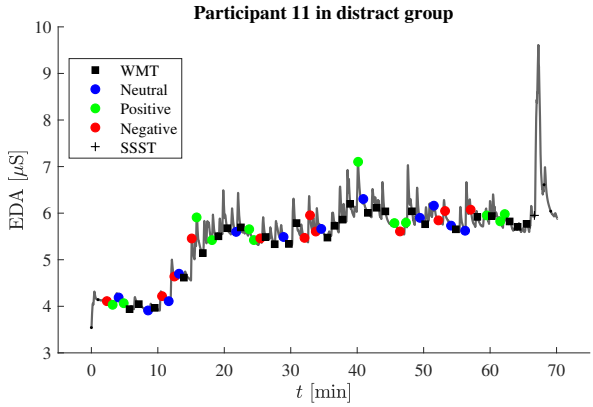

Participant 08 in attend group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

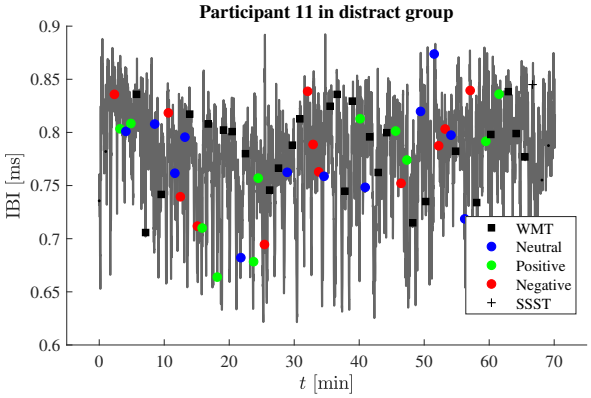

Participant 09 in distract group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

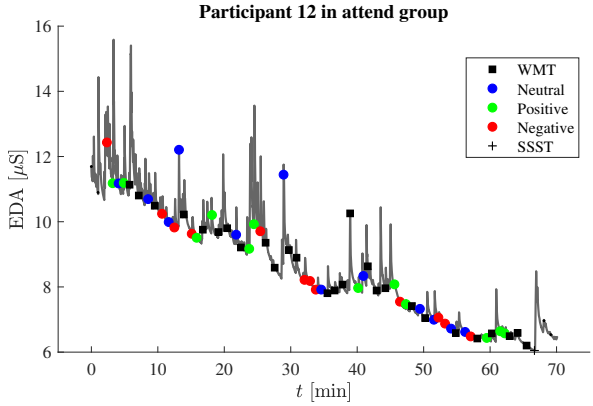

Participant 09 in distract group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

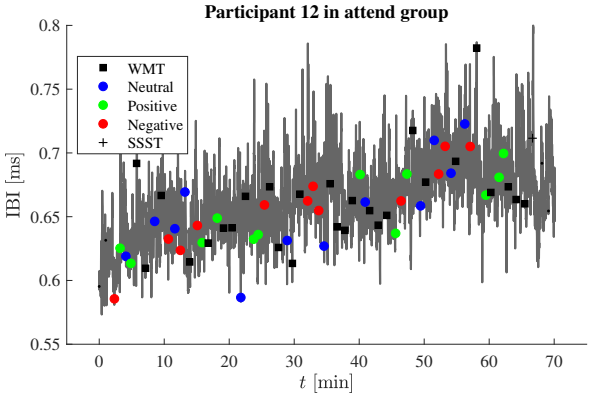

Participant 10 in attend group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

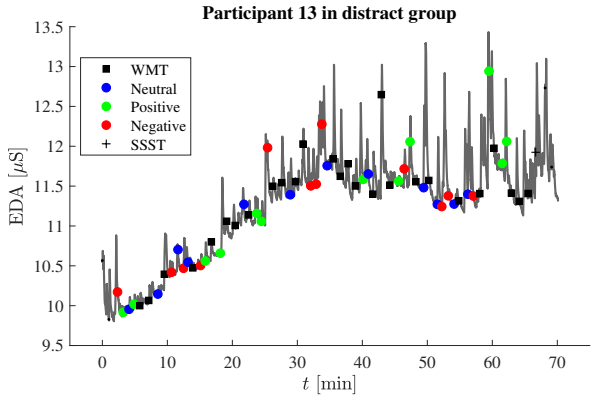

Participant 10 in attend group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

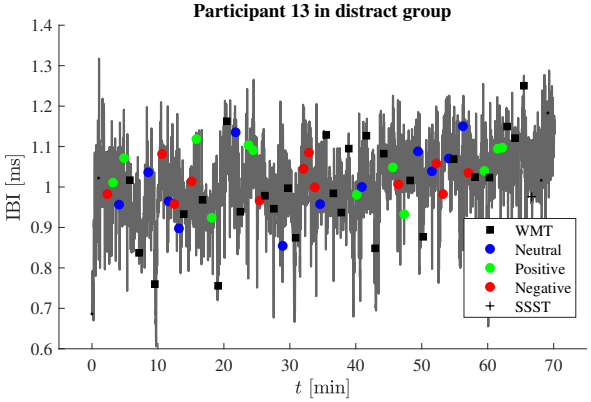

Participant 11 in distract group
WMT
Neutral
Positive
Negative
SSST
EDA [μS]
t [min]

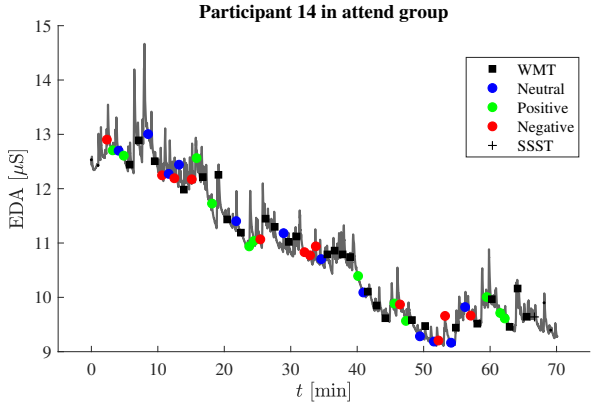

Participant 11 in distract group
WMT
Neutral
Positive
Negative
SSST
IBI [ms]
t [min]

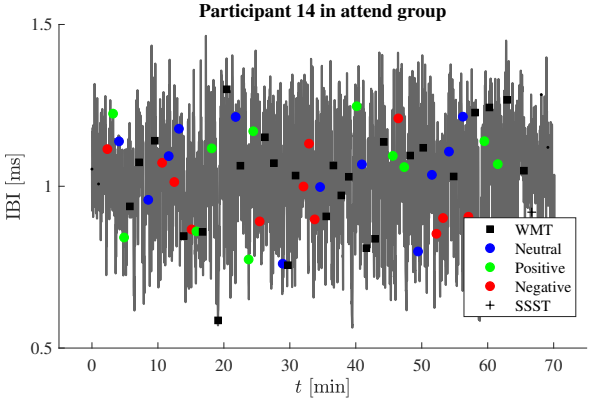

Participant 12 in attend group
WMT
Neutral
Positive
Negative
SSST
EDA [μS]
t [min]

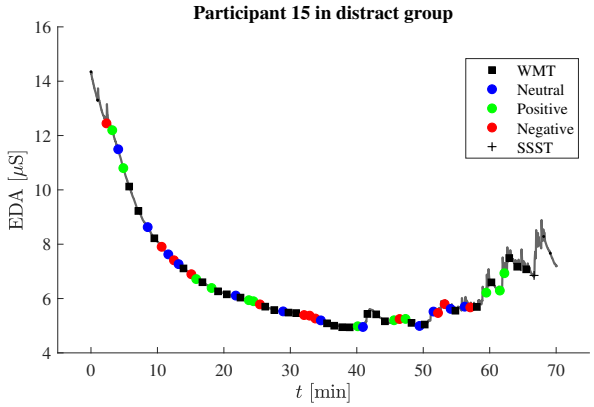

Participant 12 in attend group
WMT
Neutral
Positive
Negative
SSST
IBI [ms]
t [min]

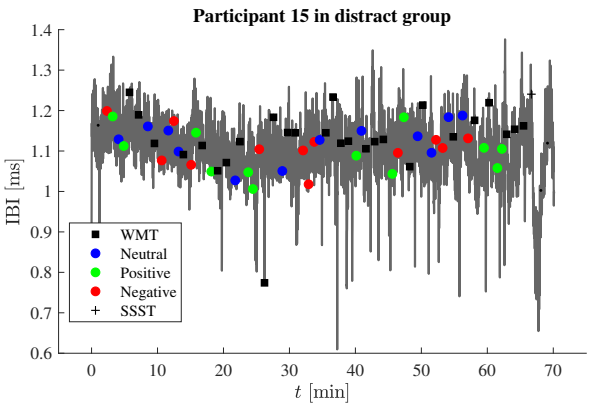

Participant 13 in distract group
WMT
Neutral
Positive
Negative
SSST
EDA [μS]
t [min]

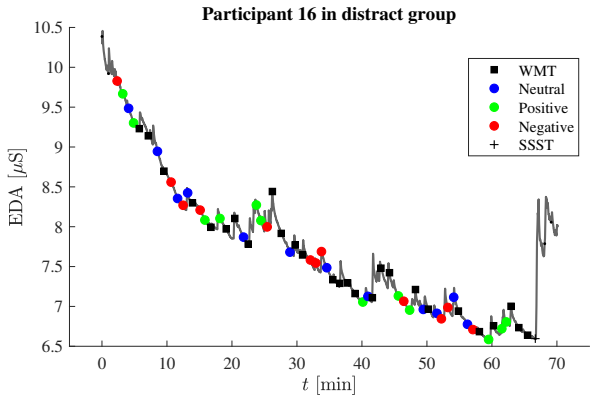

Participant 13 in distract group
WMT
Neutral
Positive
Negative
SSST
IBI [ms]
t [min]

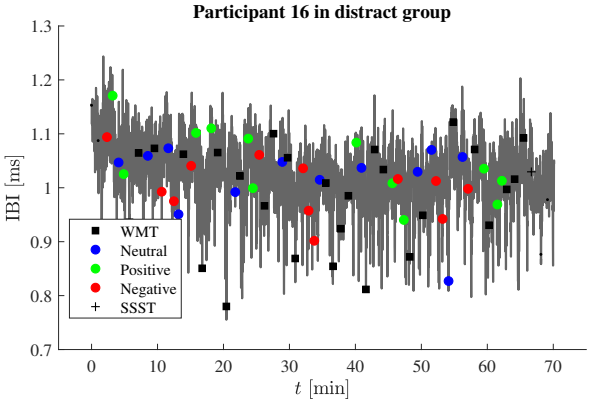

Participant 14 in attend group
WMT
Neutral
Positive
Negative
SSST
EDA [μS]
t [min]

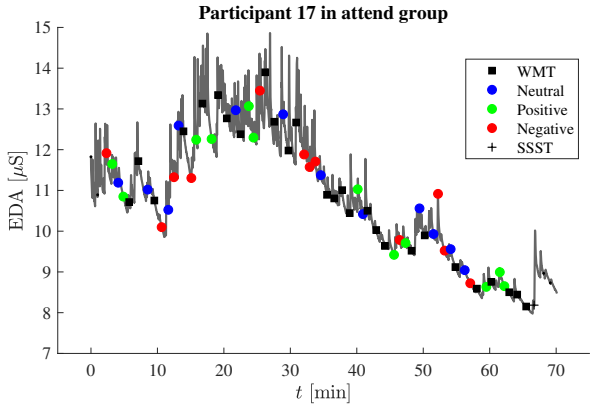

Participant 14 in attend group
WMT
Neutral
Positive
Negative
SSST
IBI [ms]
t [min]

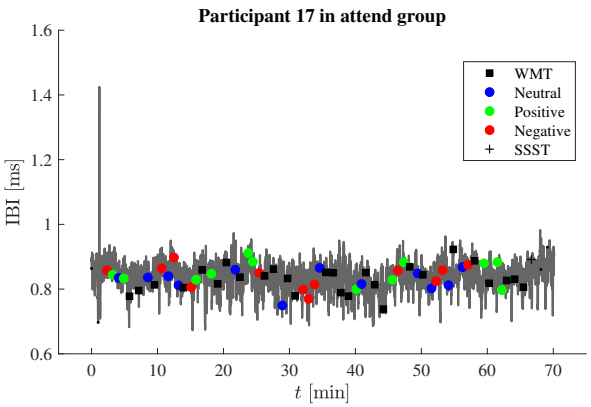

Participant 15 in distract group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

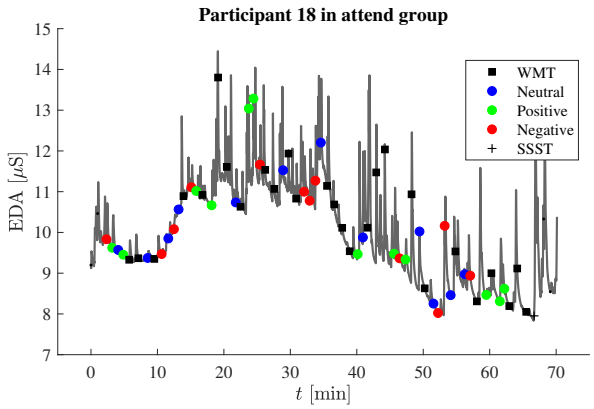

Participant 15 in distract group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

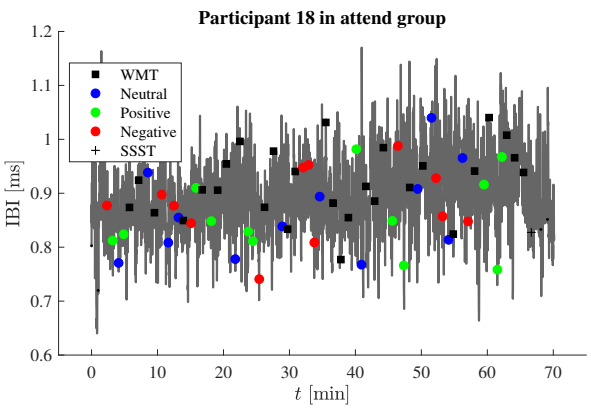

Participant 16 in distract group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

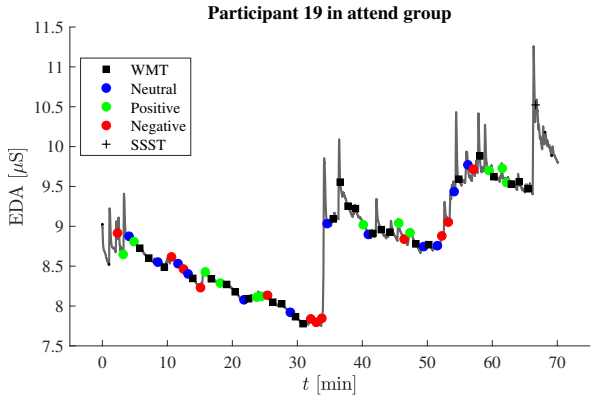

Participant 16 in distract group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

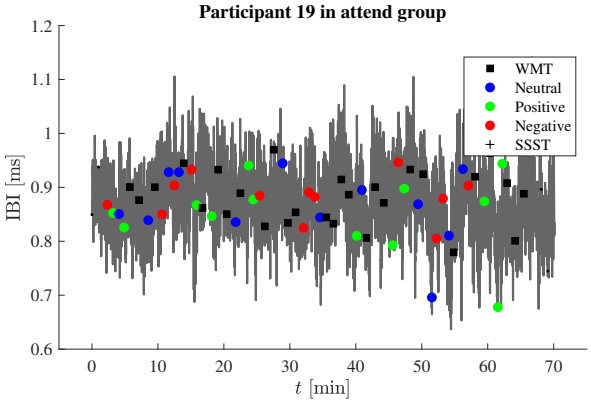

Participant 17 in attend group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

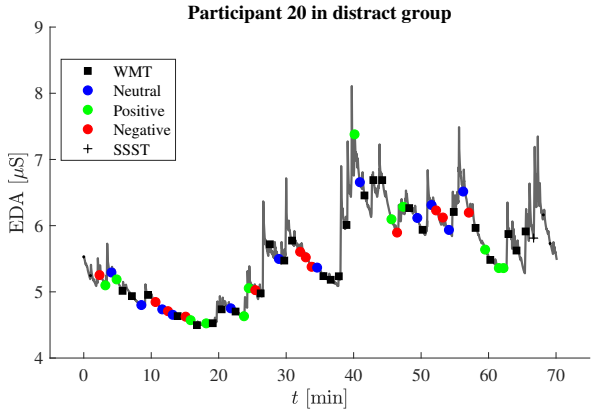

Participant 17 in attend group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

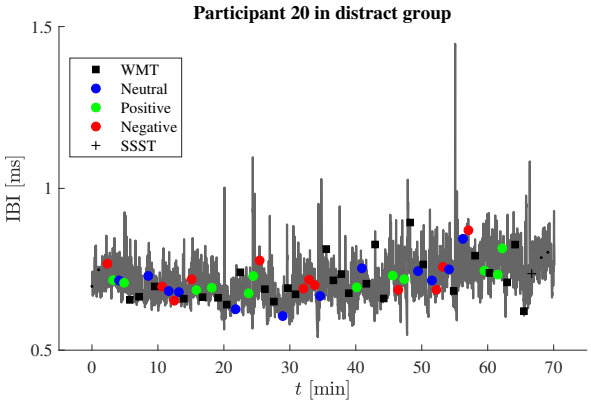

Participant 18 in attend group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST

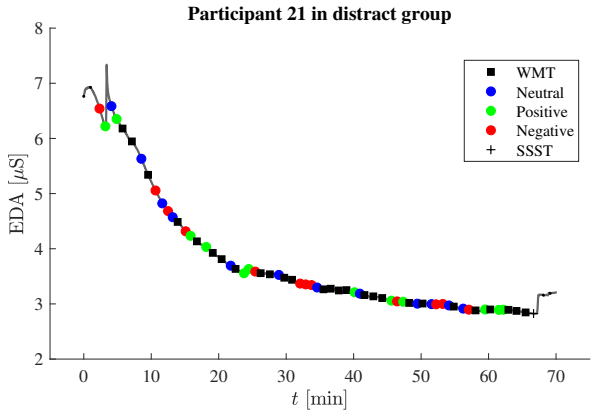

Participant 18 in attend group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

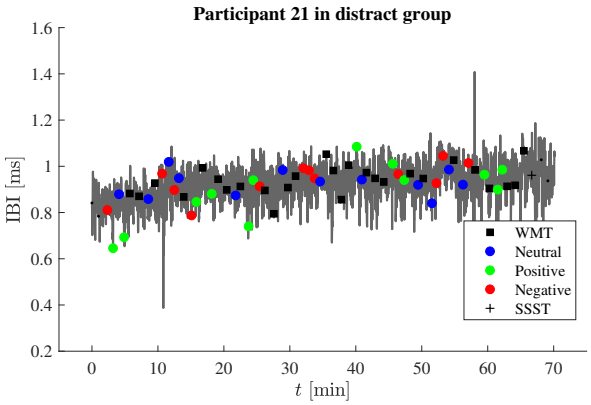

Participant 19 in attend group
WMT
Neutral
Positive
Negative
SSST
EDA [μS]
t [min]

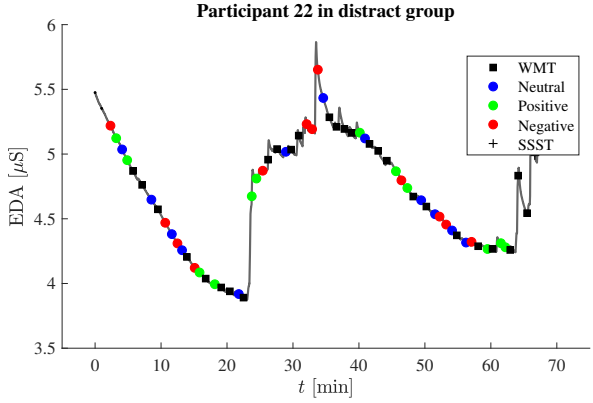

Participant 19 in attend group
WMT
Neutral
Positive
Negative
SSST
IBI [ms]
t [min]

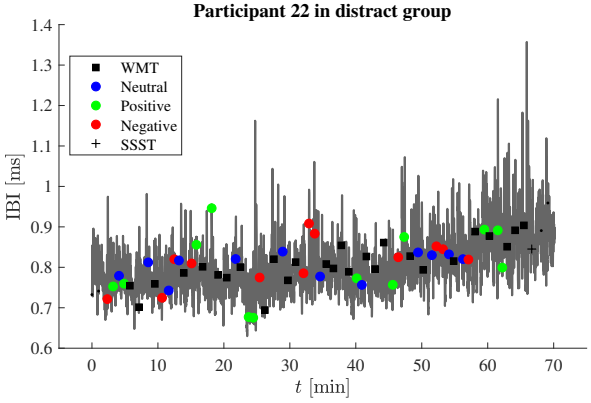

Participant 20 in distract group
WMT
Neutral
Positive
Negative
SSST
EDA [μS]
t [min]

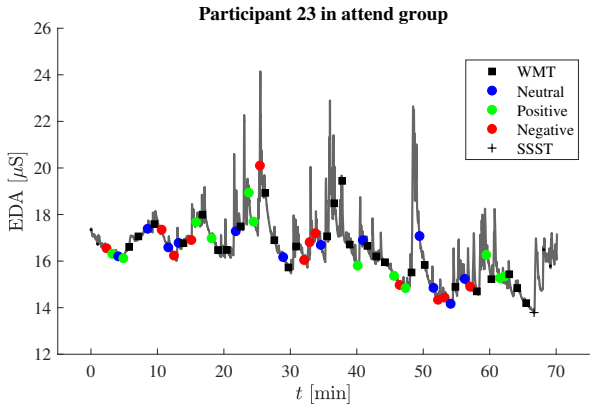

Participant 20 in distract group
WMT
Neutral
Positive
Negative
SSST
IBI [ms]
t [min]

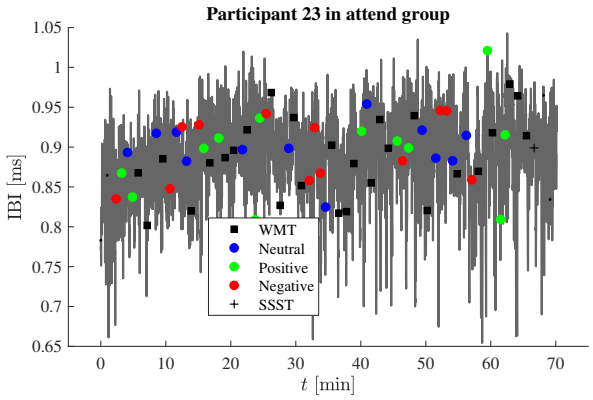

Participant 21 in distract group
WMT
Neutral
Positive
Negative
SSST
EDA [μS]
t [min]

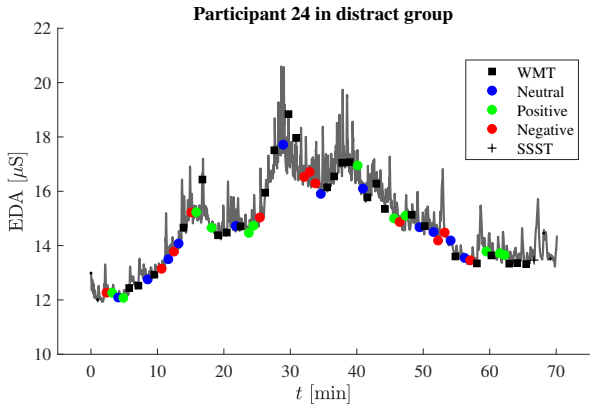

Participant 21 in distract group
WMT
Neutral
Positive
Negative
SSST
IBI [ms]
t [min]

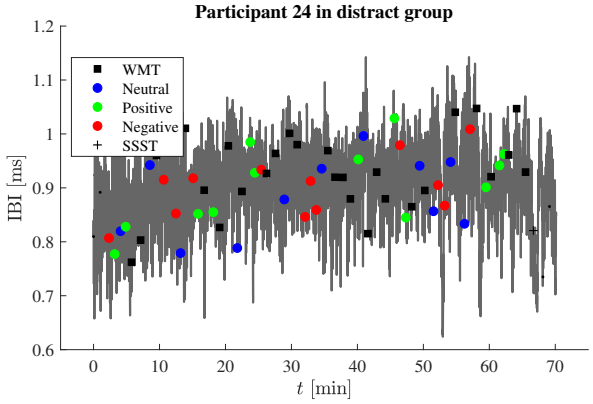

Participant 22 in distract group
WMT
Neutral
Positive
Negative
SSST
EDA [μS]
t [min]

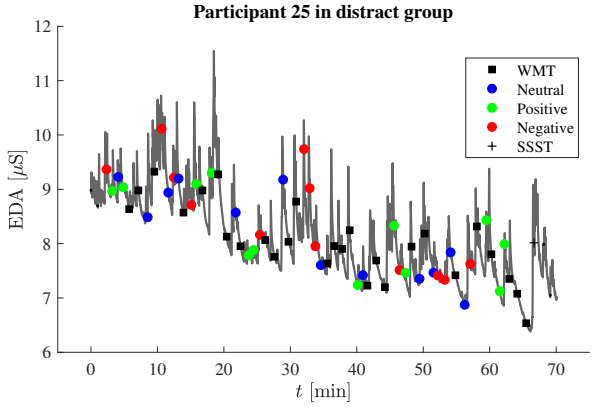

Participant 22 in distract group
WMT
Neutral
Positive
Negative
SSST
IBI [ms]
t [min]

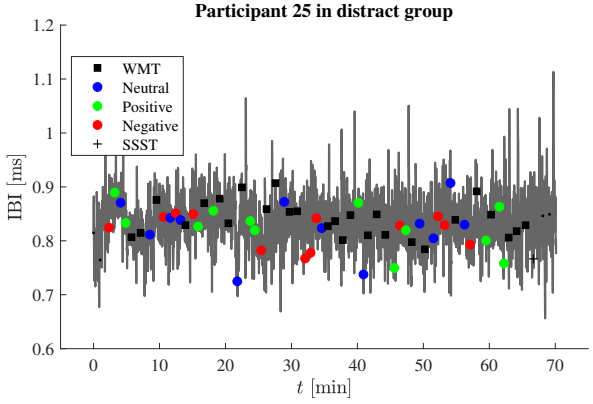

Participant 23 in attend group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST


Participant 23 in attend group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST


Participant 24 in distract group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST


Participant 24 in distract group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST


Participant 25 in distract group
EDA [μS]
t [min]
WMT
Neutral
Positive
Negative
SSST


Participant 25 in distract group
IBI [ms]
t [min]
WMT
Neutral
Positive
Negative
SSST

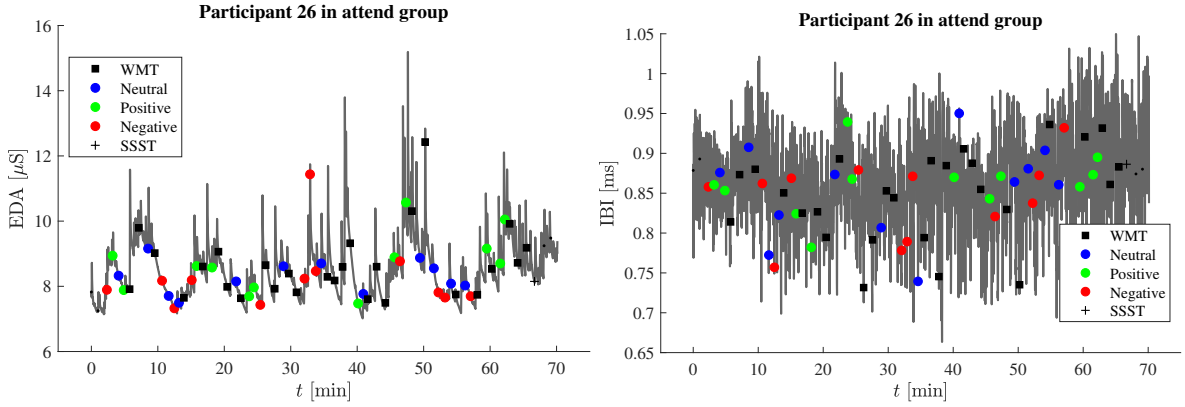


Figure F.1 Time-traces of raw electrodermal activity (EDA) and inter-beat interval (IBI) over the entire experiment for each participant. Markers show onset of working-memory task trials (WMT), affective sounds (Neutral, Positive, Negative) and the sing-a-song stress test (SSST). 'Attend' here corresponds to narrative-attending and 'distract' corresponds to short-stimuli-attending.

F.2. Working-memory-task-locked neural event-related potentials

Neural event-related potentials as induced by the WMT were not only obtained by averaging responses over each tone in a trial, but also by analyzing only responses as induced by the first tone of each task. Figure F.2 shows the grand-average event-related potentials as induced by the first tone of WMT sequences. Compared to responses averaged over all tones in a sequence (see Figure 3.1a), response amplitudes are higher toward the first tone only. Running independent-sample t-tests adjusted for multiple comparisons did not reveal significant between-group differences.

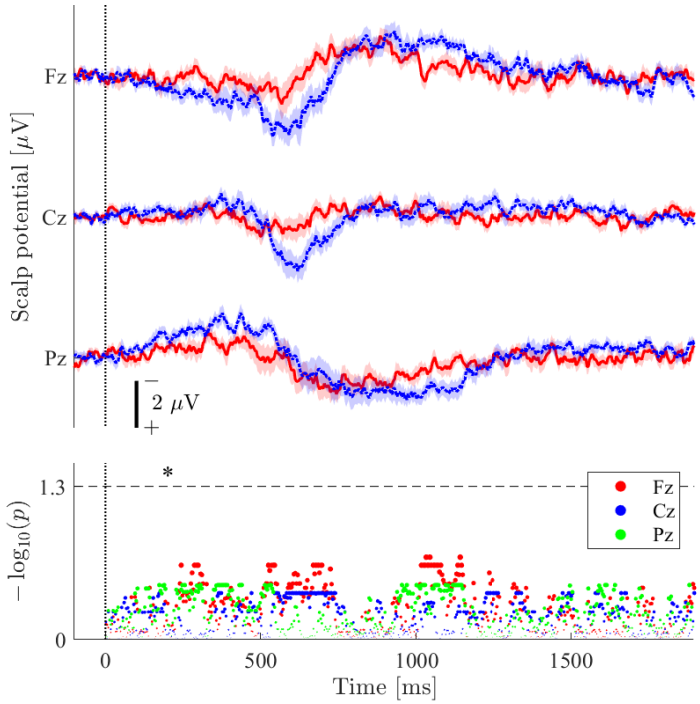


Figure F.2 Centro-parietal midline event-related potentials time-locked to stimulus onset, averaged over narrative-attending participants (red, -) and short-stimuli-attending participants (blue, - -). The standard error of the mean across participants in each group is depicted in colored areas. Significant between group differences ($p < 0.05$, multiple-comparisons adjusted independent-sample t-tests) are depicted with gray areas in the potential plots. Additionally, the corresponding adjusted p-values of between-group significance tests, including significance thresholds ($*p < .05$, $**p < .01$, $***p < .001$), are shown on a logarithmic scale below the potential plots.

F.3. Inter-subject correlations separately analyzed for the first three correlated components

Neural ISC of each of the two groups were further analyzed separately for each of the first three correlated components for the sake of comparability with earlier work. Figure F.3 shows these ISC across participants in each group. Independent-sample t-tests were conducted to test for significant between-group differences in each component, separately for each of the three (sub)sets of potentials. Test statistics can be found in Table F.1. When considering the entire narrative, no significant between-group difference were found in the first two components. In the third component, ISC were higher for narrative-attending participants. When only considering data obtained during WMT presentation, ISC in the second components were significantly higher for short-stimuli-attending participants than narrative-attending participants. For the set of data obtained during IADS-2 presentation, ISC was not

significantly different in any of the components.

The effect of short-stimulus presentation on the ISC of each group of participants was also monitored separately for each of the three components. Test statistics can be found in Table F.2. Strong significant differences were found across short-stimuli-attending participants. ISC during WMT presentation were significantly higher than ISC during the entire narrative stimulus in all three correlated components. For the IADS-2, differences were significant in the first two components. Across narrative-attending participants, effects were not as pronounced. ISC during the WMT were signifantly larger than during the entire naturalistic stimulus in the second component only ($p < .05$). During affective sound presentation, ISC were higher in the second and third component ($p < .05$).

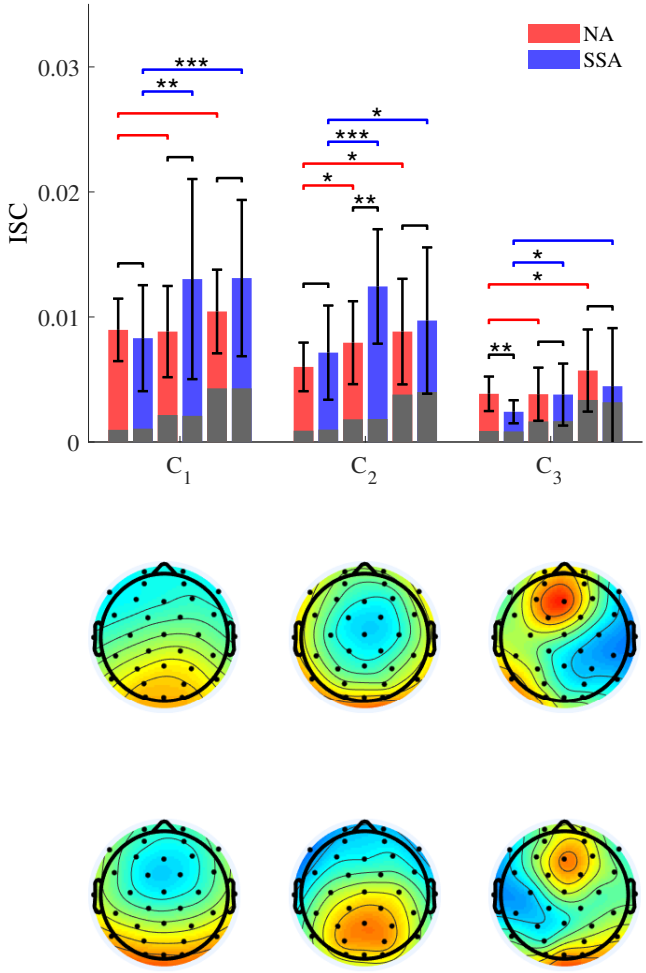


Figure F.3 Top: Mean and standard deviation of inter-subject correlations (ISC) of the first three correlated components (C_1 - C_3) across narrative-attending (red) and short-stimuli-attending participants (blue). For each components, the first two bars depict ISC of potentials across the entire naturalistic stimulus, the next two bars depict ISC across potentials obtained during working memory task presentation and the last two bars depict ISC across potentials obtained during affective sounds. Gray bars depict chance level correlations. Bottom: scalp topographies of the first three correlated components trained on narrative-attending participants (top row) and short-stimuli-attending participants (bottom row). ($*p < .05$, $**p < .01$, $***p < .001$)

F.4. Chance-level neural inter-subject correlations

Chance level ISC were determined using phase randomized surrogate data. Phase of the EEG time-series was randomized in the frequency domain, following Theiler et al. (1992). This procedure was performed separately for each of the two attentional groups and for each of the three (sub)sets of data, containing scalp potentials obtained during the entire naturalistic stimulus and during presentation of WMT trials or IADS-2 trials, respectively. Paired-sample t-tests revealed all ISC to be above chance-level, except ISC of the short-stimuli-attending group during affective sound presentation in the third component. These results are similar to earlier work, where ISC were also well above chance level in the first two components and close to chance level in the third component (Dmochowski et al., 2014, 2012; Ki et al., 2016).

Table F.1 Independent-sample t-test statistics of between-group inter-subject correlation differences for the first three correlated components (C_1 - C_3) and aggregated over these components ($\sum C_{1:3}$). Tests were conducted separately for the set of potentials obtained across the entire narrative stimulus (All) and for subsets only containing potentials during working memory task (WMT) or affective sound (IADS-2) presentation.

	C_1	C_2	C_3	$\sum C_{1:3}$
All	$t(24) = -0.49$, $p = .631$	$t(24) = 0.97$, $p = .341$	$t(24) = -3.12$, $p = .005$	$t(24) = -0.39$, $p = .698$
WMT	$t(24) = 1.72$, $p = .098$	$t(24) = 2.87$, $p = .008$	$t(24) = -0.03$, $p = .974$	$t(24) = 2.26$, $p = .033$
IADS-2	$t(24) = 1.36$, $p = .187$	$t(24) = 0.44$, $p = .662$	$t(24) - 0.79$, $p = .435$	$t(24) = 0.64$, $p = .528$

Table F.2 Paired-sample t-test statistics of between-stimulus-condition inter-subject correlation differences in narrative-attending (NA) and short-stimuli-attending (SSA) participants for the first three correlated components (C_1 - C_3) and aggregated over these components ($\sum C_{1:3}$).

			NA	
	C_1	C_2	C_3	$\sum C_{1:3}$
WMT vs. All	$t(12) = -0.30$, $p = .773$	$t(12) = 2.71$, $p = .019$	$t(12) = -0.05$, $p = .960$	$t(12) = 1.38$, $p = .193$
IADS-2 vs. All	$t(12) = 1.15$, $p = .273$	$t(12) = 2.50$, $p = .028$	$t(12) = 2.33$, $p = .038$	$t(12) = 3.77$, $p = .003$

			SSA	
	C_1	C_2	C_3	$\sum C_{1:3}$
WMT vs. All	$t(12) = 3.97$, $p = .002$	$t(12) = 6.50$, $p < .001$	$t(12) = 2.22$, $p = .046$	$t(12) = 8.11$, $p < .001$
IADS-2 vs. All	$t(12) = 4.33$, $p < .001$	$t(12) = 2.32$, $p = .038$	$t(12) = 1.68$, $p = .118)$	$t(12) = 4.17$, $p = .001$

Table F.3 Test statistics of paired-sample t-tests conducted to test whether inter-subject correlations were higher than chance-level correlations in each of the three components (C_1 - C_3) and summed over these components ($\sum C_{1:3}$) for (sub)sets of data containing scalp potentials obtained during the entire naturalistic stimulus (All) and during presentation of working-memory task (WMT) trials or affective sound (IADS-2) trials for both narrative-attending (NA) and short-stimuli-attending (SSA) participants.

	NA			
	C_1	C_2	C_3	$\sum C_{1:3}$
All	$t(12) = 12.12$, $p < 10^{-7}$	$t(12) = 9.21$, $p < 10^{-6}$	$t(12) = 7.13$, $p < 10^{-4}$	$t(12) = 13.43$, $p < 10^{-7}$
WMT	$t(12) = 6.28$, $p < 10^{-4}$	$t(12) = 6.83$, $p < 10^{-4}$	$t(12) = 3.58$, $p = .004$	$t(12) = 7.37$, $p < 10^{-5}$
IADS-2	$t(12) = 6.93$, $p < 10^{-4}$	$t(12) = 4.84$, $p < .001$	$t(12) = 2.54$, $p = .026$	$t(12) = 10.57$, $p < 10^{-6}$

	SSA			
	C_1	C_2	C_3	$\sum C_{1:3}$
All	$t(12) = 5.87$, $p < 10^{-4}$	$t(12) = 5.57$, $p < .001$	$t(12) = 5.93$, $p < 10^{-4}$	$t(12) = 6.52$, $p < 10^{-4}$
WMT	$t(12) = 4.83$, $p < .001$	$t(12) = 8.64$, $p < 10^{-5}$	$t(12) = 3.19$, $p = .008$	$t(12) = 7.03$, $p < 10^{-4}$
IADS-2	$t(12) = 4.94$, $p < .001$	$t(12) = 3.21$, $p = .008$	$t(12) = 0.95$, $p = .361$	$t(12) = 4.73$, $p < .001$