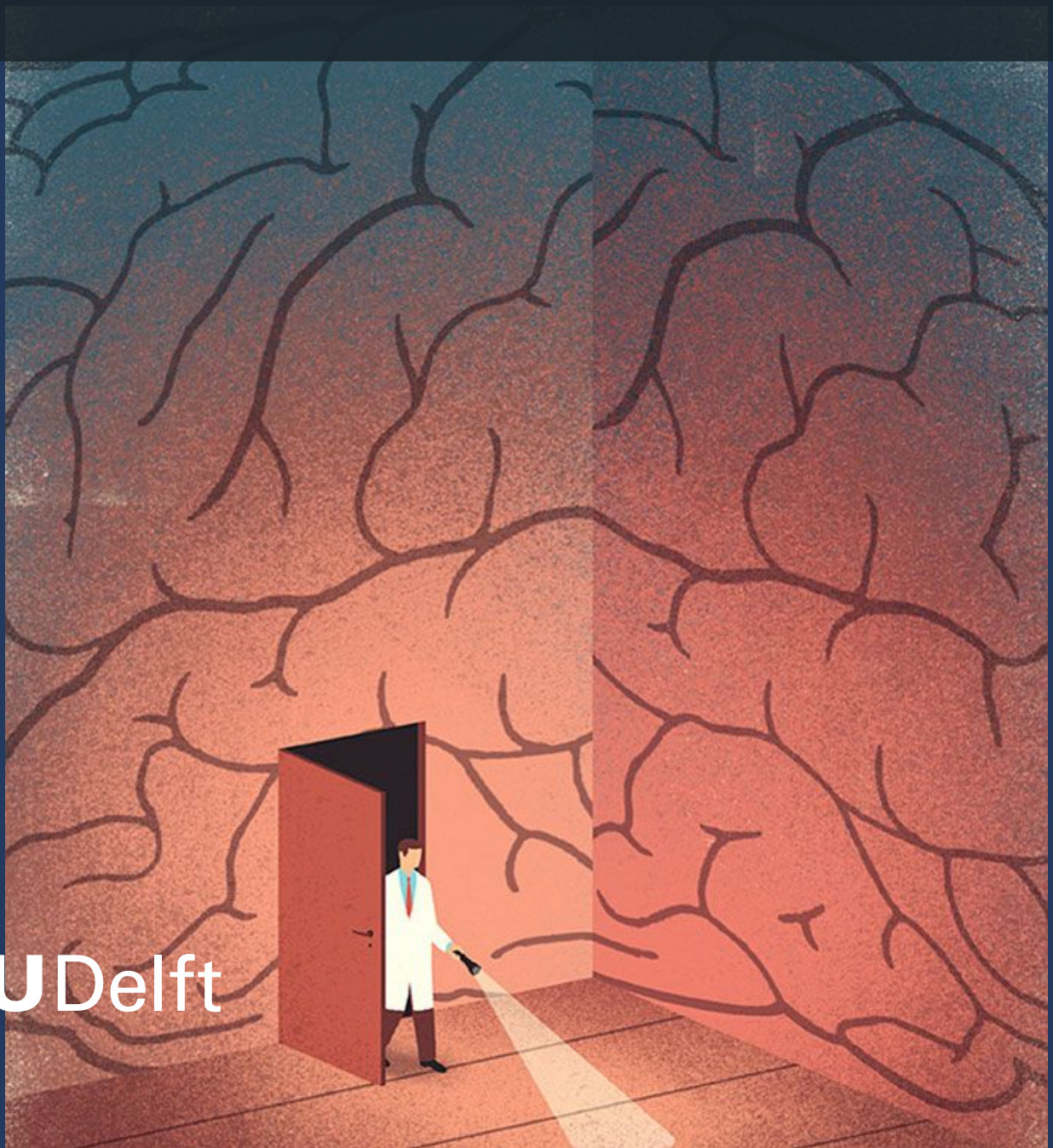
Physiological Sensor Suite: Design and Implementation

Measuring Startle & Surprise using EEG, EMG and ECG

Ward Bogaerts



TU Delft

Physiological Sensor Suite: Design and Implementation

Measuring Startle & Surprise using EEG, EMG and ECG

Thesis report

by

Ward Bogaerts

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended publicly on October 26, 2023 at 12:45

Thesis committee:

Chair:	Prof. dr. ir. M. Mulder
Supervisors:	Dr. ir. M.M. van Paassen
	Ir. O. Stroosma
	Dr. A.H. Landman
External examiner:	Dr. ir. D.M. Pool
	Dr. A. Bombelli
Place:	Faculty of Aerospace Engineering, Delft
Project Duration:	December, 2022 - October, 2023
Student number:	5378273

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

Faculty of Aerospace Engineering · Delft University of Technology

Preface

Welcome to the culmination of nine months full of blood, sweat and tears. I am hereby proud to present my final thesis. Starting on this topic was like learning to ride a bike all over again. Trying, falling, and getting back up. However, this was the reason that attracted me to this topic in the first place: a completely unknown territory of knowledge. You can't learn to ride a bike on your own, so that is why I have to thank a few people.

Starting off, I would like to express my gratitude towards Max Mulder, who proposed me this topic and gave me a lot (too much?) freedom from start to finish. Without him, this document would have never existed in the first place. Thank you for trusting me, both with your time, and with the university's money. I secretly hope that I can tag along on your next trip to Japan, arigato gozaimasu.

Second, I would like to thank my daily supervisor, and DUECA wizard, René van Paassen. Without him, I would have thrown my computer in one of the Delft canals after yet another DUECA programming mistake. In addition, René was always the first person with whom I could discuss new ideas, arising problems and possible solutions. He guided me throughout the whole process and was always available to help out when necessary. I hope you get to enjoy canoeing very soon.

I would like to thank Olaf Stroosma for reminding me about all the practical aspect of performing an experiment. Without him, I would probably have been performing my experiments in the entrance hall of the faculty without any preparation. Thank you for assisting both with the experiment, recruiting participants, and helping out with DUECA while always being straight to the point. Good luck with teaching the next generation about the wonderful world of DUECA.

Many thanks to Annemarie Landman for her expert advice regarding startle & surprise. It was immediately clear that you are the go-to person in this field, which helped a lot in orienting during the literature study. Without you, I would have been lost in the depths of the neurophysiology journals. In addition, thank you for helping out with the experiment statistics. May results always be significant for you.

I would also like to thank my family, from providing mental support to actually participating in the experiment. Thank you for supporting me throughout this long study journey and never failing to believe in me. Thank you Kato for being my patient zero during the experiment development and for taking up the role of independent rater during your spare time. Lastly, thanks to my neighbours Els and Mirte for travelling all the way to Delft, just to get startled.

Hope you enjoy reading,

Ward Bogaerts

Delft, 26 October 2023

Contents

List of abbreviations

DUECA Delft University Environment for Communication and Activation. 99

ECG electrocardiography. v, 99, 104, 107
EEG electroencephalography. v, 99, 104, 108–113
EMG electromyography. v, 99, 104–106
ERP event-related potential. 104

HR heart rate. 104, 107

ICA independent component analysis. v, 104, 108–110

List of Figures

List of Tables

Part I

Scientific Article

Physiological Sensor Suite: Detecting Startle & Surprise Using EEG, EMG and ECG

Ward Bogaerts
Faculty of Aerospace Engineering
Delft University of Technology
Delft, the Netherlands
w.bogaerts@student.tudelft.nl

***Abstract*—Today, most fatal accidents in commercial aviation are caused by loss of control in flight. Due to the increased reliance on automation and pilots being moved out of the loop, startling and surprising events can be contributing factors. To more effectively train pilots, startle and surprise are to be included in training scenario's. A subjective scale indicating the level of startle and surprise is being developed to help build these training programmes, but an objective baseline of whether pilots are startled or surprised is needed in order to validate and further develop this scale. In this way, the scale can be used in flight simulators to asses the pilot's perceived level of startle and surprise during novel training scenario's. This work aims to test a startle and surprise detection method using physiological data. Electromyography, electrocardiography and electroencephalography data were collected. A validation of the occurrence of physiological effects related to both startle and surprise was performed, after which a detection algorithm was constructed. Validation of effects was performed on data of 22 participants using a three-stimulus oddball task with additional auditory startling stimuli. It was found that all considered physiological effects related to startle can be reliably observed. Contrary, not all physiological effects related to surprise were observed. The detection algorithm was tuned on data of twelve participants and showed to generalise well to the other ten data points. Startle detection could be performed with high accuracy, although surprise detection was poor.**

***Index Terms*—startle, surprise, aviation, EMG, ECG, EEG, detection, physiological effect, independent component analysis, ICA, P300, ERP, event-related potential**

I. Introduction

Loss of control still forms the largest category of fatal accidents in commercial aviation [1]. With increased levels of automation autonomy and authority in the flight deck, supervisory control has become the dominant role of the pilot. However, in situations that can not be handled by automated systems, the role of the pilot becomes critical and manual intervention may be needed. These situations tend to be unforeseen and complex, inducing the need for rapid adaptation to the situation and quick judgement and decision making [2]. The switch to an active role can be difficult however after e.g. long periods of automated flight [3]. Furthermore, pilots' manual flying skills degrade with extensive use of automation. In addition, over-reliance on the automated system can decrease the level of active monitoring by the pilot. Due to the automation not being completely transparent, pilots can not obtain full knowledge of the process executed by the automation, which can lead to automation surprises. [4]

Because the pilot has been moved out of the loop, a startle factor can arise during unexpected situations. Startle is defined as a fast response to sudden, intense stimuli to protect the organism from harm [5]. This startling effect often complicates the crew's ability to troubleshoot the problem [6], [7]. Surprise is defined as an emotional response to something unexpected, resulting from a mismatch between the expectations and perception of the environment [8].

To reduce the issue that startling and surprising effects impose, two research directions are being explored. The first deals with preventing the problem by increasing the quality of the automated systems. An example of research in this direction is the development of ecological interfaces to give the pilot insight into the physics behind the automation [9]. Other examples include already implemented automated safety systems such as conflict detection and resolution systems [10]. If, however, surprising or startling situations still occur, the pilot profits from having the appropriate skills necessary to deal with the situation. This is the main motive for the second research direction: improving pilot resilience in startle & surprise (S&S) situations through training [11]–[13].

A conceptual model of S&S has been proposed by Landman *et al.* [14] to explain pilot performance in surprising and startling situations. In addition, a subjective rating scale to indicate S&S, similar to the NASA TLX scale for measuring workload [15], is being developed.

The main goal of this work is to test a S&S sensor suite developed to automatically detect S&S using physiological data. This system can then assist in validating the subjective scale by introducing objective quantitative data and validating the subjective data. Having the scale allows for tuning improved training scenarios that also include S&S. In addition, the implemented sensors can aid in validating non-invasive methods to monitor the pilot's state during flight [16].

To find whether S&S can be reliably detected, an experiment was designed to specifically induce S&S in participants. During the experiment, physiological data were collected to be used for detecting S&S. This was done in two phases. The first phase consisted of validating whether the physiological effects related to S&S occur and could be observed using a physiological data acquisition system. This was followed by

the construction of a detection algorithm which automatically detects the occurrence of these physiological effects after stimulus onset and classifies a certain stimulus as surprising, startling, or neither.

Due to the startle reaction being highly physiological [17], [18], we expect clear and reliable startle indicators which can be used to detect the startle reaction automatically. Since surprise is a more cognitive process and ambiguous results are reported in literature [8], [19], [20], we expect greater difficulty in finding reliable indicators.

The remainder of this paper is structured as follows: Section II provides background information about S&S and the used measurement techniques. Section III describes the experiment and method used for the experiment and validation of physiological effects. After, Section IV will describe the results obtained from the validation of the occurrence of the physiological effects. The method used for constructing the detection algorithm is described in Section V. Results from the detection algorithm will be presented in section VI. A discussion about the results will be presented in Section VII, where the results will be compared against the hypotheses as well. Section VIII will provide conclusions about the main findings of this work.

II. Background

To accurately detect S&S, we must identify the physiological effects associated with S&S and select those that can be measured in a flight setting. This section provides background information about S&S and their physiological effects. After, a selection of measurements is described and earlier implementations of these measurements in flight or car settings are provided.

A. Startle & surprise

As mentioned in the introduction, startle is defined as a fast response to sudden, intense stimuli to protect the organism from harm [5]. It generates a fast, automatic and often highly physiological reaction called the startle reflex, combined with a slower, conditioned and behavioural reaction called the startle response, where focus shifts towards the source of the startling stimulus [17], [18]. The startle reaction is increased by unexpectedness. However, anticipated stimuli can be startling as well [14]. An example of a startling event would be a lightning strike during a storm. Startling events do not have to be task-relevant [17].

As mentioned, surprise is defined as an emotional response to something unexpected, resulting from a mismatch between the expectations and perception of the environment [8]. This causes a mismatch between the current mental scheme and the observed reality, called a schema-discrepancy [8], [14], [21]. It forces re-evaluation of the situation and changing a person's understanding. Surprise causes the interruption of the person's current task and induces an orienting attention shift to the cause of the surprise [14], [17]. While startle is always a response to a certain stimulus, surprise can also be the consequence of the absence of a stimulus [17]. An example would be an odd event such as a dial indicating a unexpected value. Contrary to startle, surprise is mostly elicited by the presence (or absence) of a task-relevant stimulus [14].

B. Physiological effects of startle

The startle reaction is composed of two separate responses: the initial motor response (also called muscular tension reflex or startle reflex), and the orienting response (also called startle response) [22]. The initial motor response is fast and has an onset latency of 20-50 ms [22]. It is characterised by muscle contraction in a "fight-or-flight" reaction to the stimulus. No control over this response can be achieved, it is purely reflexive. The response starts with eye blinking, and is followed by a forward head movement, facial grimacing, shoulder elevation, arm abduction, elbow bending, forearm pronation, finger flexion, abdominal contraction and knee flexion [17], [23]. Most muscle activity is located in the face and shoulders, notably the blinking reflex and tension of the trapezius muscles [14], [18], [22].

The orienting response is used to focus attention on the startling stimulus and has an onset latency of 400-450 ms [22]. It is determined by the stimulus' nature, causing a source of variance [24]. Contrary to the initial motor response, the orienting response is an emotional and voluntary response [22]. It involves orienting to the stimulus source, increased skin conductance, increased systolic blood pressure (BP), heart rate (HR) and pupil size [18], [22], [23]. The initial motor response is valence-related: increased during negative emotional states and decreased during positive emotional states [25]. The orienting response is not valence-related [26].

C. Physiological effects of surprise

The surprise reaction, much like the startle response, consists of two phases: the response to the unexpectedness of the event, and the sense-making of this event [21]. Response to unexpectedness is independent of the valence of the surprising stimulus, while the sense-making can invoke different psychophysiological reactions depending on its valence [21]. There is much discussion on the physiological effects of surprise [8], [19], [20]. This induces the need for validating the effects at hand.

Considering the reaction to unexpectedness, an increase in the electroencephalography (EEG) P300 event-related potential (ERP) level can be observed, with larger P300 values for more schema-discrepant events [21], [27], [28]. The P300 response is combined with an increased activity in the delta (1-3Hz) and theta (3-6Hz) EEG frequency bands [29], [30]. The P300 ERP originates in the anterior cingulate cortex, which shows an increase in activity around 300 ms after the schema-discrepancy. This can be observed by a positive deflection in the EEG [31]. The P300 ERP is unrelated to the valence of the stimulus [21]. In addition, unexpectedness increases pupil dilation [8], [32]. However, the specificity of pupil dilation is still under debate [8].

This is followed by attentional allocation and interruption of ongoing behaviours, which is characterised by action

delays, increased error rate and surprise induced situation blindness occurring as early as 200 ms after the schema-discrepancy is noticed [8], [21]. Specific reactions include gaze towards the stimulus, occurring around 400 ms after discrepancy [8], [33]. In addition, a surprise expression can be observed, constructed of raised eyebrows, widened eyes and open mouth [21]. However, except for increased corrugator activity, Reisenzein, Bördgen, Holtbernd, *et al.* [34] argue that these effects have not been reproduced sufficiently. An initial HR deceleration lasting for around one second can be observed after the surprising stimulus [35], [36]. Between the reaction to unexpectedness and sense-making, verification of the schema-discrepancy is performed by taking a second look at the surprising event [8].

Sense-making is associated with cardiovascular reactivity and depends on the nature of the event [21]. It is accompanied by an increased HR starting at around one second after stimulus onset and lasting for about three seconds [35], [37].

D. Selection of measurements

Three physiological measuring methods were chosen: electromyography (EMG), electrocardiography (ECG) and EEG. These methods allow the detection of physiological effects summarised in Table I, including their onset latency and measurement method. Using three measurement techniques, eight physiological effects can be measured. In this report, we count the decrease and subsequent increase in HR after a surprising stimulus as a single effect, called the "surprise HR pattern". The measurements cover the most constant and sensitive parameters which can be measured , apart from using eye tracking, along with less constant parameters for which supporting measurements are required.

The aforementioned trapezius muscle activity was previously measured using EMG in a real world driving task [38]. Tichon *et al.* [39] used EMG in a fixed-base flight simulator. Implementing EMG for eye blink measurement has been described extensively by Blumenthal *et al.* [40]. Ryu and Myung [41] used EEG measures for a manual control task. Rooseleer *et al.* [42] uses an off-the-shelf EEG recorder in a moving-base flight simulator. Gibson *et al.* [43] used dry-electrode EEG in a moving-base flight simulator. EEG in a fixed base simulator was also performed by Johnson *et al.* [44]. Bruna *et al.* [45] implemented ECG measurements in a fixed base flight simulator, while Hannula *et al.* [46] performed the same measurements in a moving base simulator. Healey and Picard [38] implemented ECG measures in a real world driving task, while Veltman [47] performed ECG recordings in both simulated and real flight.

III. Phase I method: validation of S&S effects

The experiment served to answer the following research question:

- Do the effects listed in Table I occur when participants are subjected to surprising and startling stimuli?

Physiological data obtained from the experiment were used to observe changes in physiological effects following a stimulus when its type is known. This analysis helps pinpoint reliable indicators of surprise or startle. Section IV will deal with the results related to the validation of the occurrence of the physiological effects listed in Table I. Construction of the detection algorithm and its results will be treated in Sections V and VI.

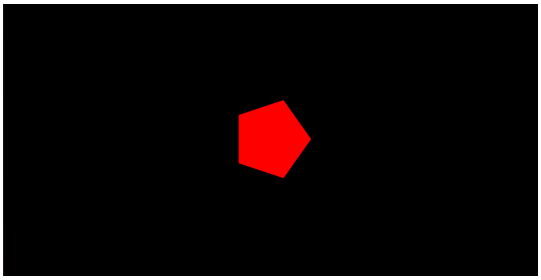
TABLE I: Physiological effects belonging to either startle or surprise

Startle		
Effect	Latency	Method
Eye blink	20-50 ms	EMG, EEG
Increased trapezius muscle activity	50-400 ms	EMG
Increased HR	3-4 s	ECG
Surprise		
Effect	Latency	Method
P300 ERP increase	300 ms	EEG
Increased delta frequency band activity	0-500 ms	EEG
Increased theta frequency band activity	250-500 ms	EEG
Decreased HR	until 1 s	ECG
Increased HR	1-4 s	ECG

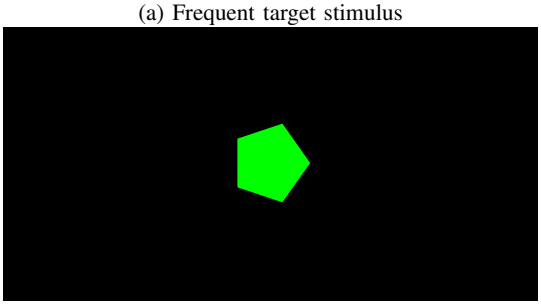
A. Stimuli and apparatus

For eliciting surprise, a three stimulus oddball paradigm, similar to [35] was used. The stimulus train consisted of three types of stimuli: frequent non-targets, frequent targets and rare non-targets. 72 stimuli were presented in a stimulus train. The ISI varied between 12, 16, and 20 seconds with equal probability. Long ISIs were chosen due to the expectancy of reliable ERPs with few enough stimuli to limit possible habituation effects [35]. The train consisted of 33 frequent targets and frequent non-targets, and six rare non-targets. This approach differs from the approach in [35] since the number of frequent targets and non-targets is equal. The rationale behind this choice is to make the frequent targets of equal valence as the frequent non-targets. Since the existence of rare non-target stimuli was unknown by participants, it is expected to result in a surprise reaction only being elicited for a rare non-target stimulus, and not for frequent target stimuli. The pentagon-shaped stimuli (in the following referred to as 'dots') only differed in colour. Frequent non-target stimuli were green (RGB(0,1,0)), frequent targets red (RGB(1,0,0)) and rare non-targets cyan (RGB(0,1,1)). Visual stimuli were presented on a computer screen with a black background and were presented for 100 ms. Figure 1 shows the different stimuli.

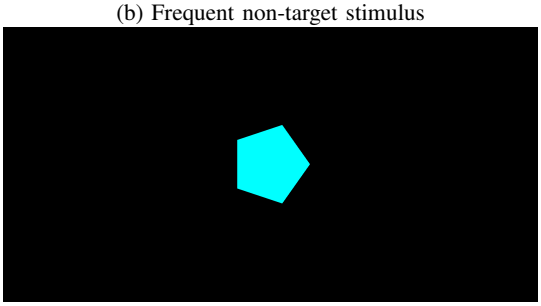
For eliciting startle, a single auditory stimulus was used, presented six times during the surprise stimulus train. The stimulus consisted of a loud gunshot sound. Startle ISIs varied between 140, 180 and 200 seconds with equal probability. Stimuli were outputted via a Logitech stereo 5.0 speaker system. The volume had an average of 93.2 dBA sound pressure level (SPL) (SD = 0.82 dBA) between stimuli. This volume range was chosen since auditory startling effects occur



(a) Frequent target stimulus



(b) Frequent non-target stimulus



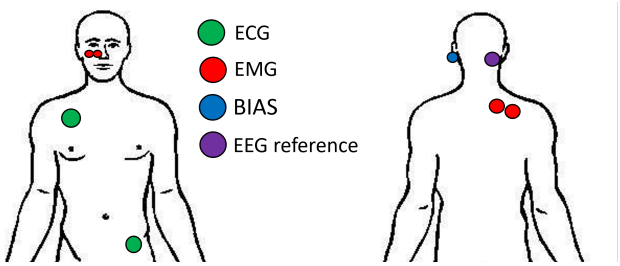
(c) Rare non-target stimulus

Fig. 1: Surprise stimuli as presented to the participant

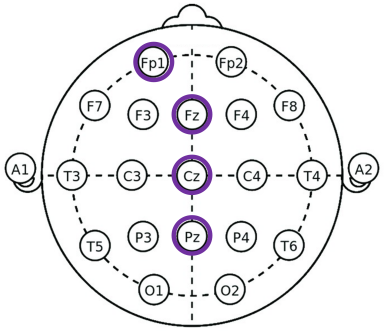
when the volume level rises above 90dBA SPL [23]. Stimulus train construction is described in Appendix A.

During the task, EMG, ECG and EEG measurements were obtained. The physiological measurements were performed using an OpenBCI Cyton board [48]. The OpenBCI Cyton board was controlled using the Brainflow application programming interface (API) inside the DUECA environment. This was done to ensure synchronisation between the Cyton data acquisition module and the modules controlling the experiment and stimulus presentation, also implemented in DUECA.

EMG measurements were performed on the right trapezius and lower right orbicularis oculi muscles. ECG leads were placed in an Einthoven Lead II configuration [49]. A dry clip-on electrode was used on the left earlobe to function as a ground signal for EMG, ECG and EEG. EEG leads were placed on the Fz, Cz, Pz and Fp1 locations, according to the 10-20 system for electrode placement [50]. The Fz, Cz, and Pz locations are most suitable for detecting the P300 ERP [51], while the Fp1 location can be used to detect eye blinks [52]. A right mastoid referene was used for the EEG measurements to remove background noise. The diagram of the placed electrodes can be seen in Figure 2.



(a) EMG, ECG and EEG electrode placement



(b) 10-20 EEG electrode placement. Figure modified from [50]

Fig. 2: Electrode placement diagram

The software used for statistical analysis and for constructing the detection algorithm was Python 3.6, combined with the NumPy, BrainFlow and Neurokit toolboxes. Detailed information about the Cyton board and electrodes can be found in Appendix B.

B. Participants & procedure

A total of 23 participants with an average age of 27.3 years (SD = 8.1 years) participated in the experiment. Participation was voluntary. The experiment lasted about 45 minutes per participant, of which 20 minutes were spent performing the test run. Participants were seated behind a desk and were looking at the display. The experiment was performed in a darkened room to minimise outside effects. Figure 3 shows the experiment set-up. The experiment was approved by the Human Research Ethics Committee of TU Delft with reference number 3203.

1) Briefing: All participants received an information letter and signed an informed consent form. The task was then again explained as follows. Participants were informed about the existence of green and red dots. They were instructed to keep a mental count of the amount of red dots that appeared throughout the experiment and to ignore the green dots.

2) Sensor placement: Briefing was followed by placing the electrodes at the locations described in Section III-A. Before placing electrodes on the trapezius or chest, participants were informed that they were free to attach these themselves if they felt uncomfortable about the researcher placing them. With the electrodes attached, the OpenBCI GUI was used to check

Fig. 3: Experiment set-up

whether all channels were outputting data. In addition, the impedance of the dry EEG electrodes was checked. If values were below $750k\Omega$, the impedance was seen as acceptable. See Table XII in Appendix C for the obtained impedances.

3) Performed counting tasks: After sensors were applied and their impedance checked, participants had to perform a training session or around 5 minutes. This training session was followed by a 20-minute test session where the participant had the same counting task.

4) Debriefing: After completing the test session, participants were asked to fill in a short questionnaire about their subjective level of startle and surprise, as described in Section III-E.

C. Training session

in the training session, ten stimuli were presented without startling or surprising stimuli. During training, a separate surprise stimulus train was used. The training session took approximately five minutes. The goal of this session was to get the participant used to both the counting task and the equipment. Furthermore, since at the beginning of the experiment, all stimuli are novel to the participant (even the frequent target/non-target stimuli), a surprise reaction can be elicited by these first stimuli as well [35]. These effects will then not occur in the post-training session since the participant will be familiar with the stimuli. During the training session, participants were allowed to ask questions and make themselves comfortable. The data during the training session were stored but not used for analysis.

D. Test session

The test session included the surprising and startling stimuli in addition to the frequent targets and non-targets. This session took approximately 20 minutes and contained 78 stimuli, of which 6 were startle, 6 rare non-targets (surprise stimuli; blue dots), 33 frequent non-targets (green dots) and 33 frequent targets (red dots). No startling stimuli were presented within three seconds of a visual stimulus. In addition, the first 10 stimuli were frequent targets or frequent non-targets only. For the test sessions, all nine remaining combinations of startle- and surprise stimulus trains were used, whereby a certain participant got assigned one combination. Participants were instructed to remain silent and reduce movement to a minimum during this session.

E. Post-experiment questionnaire

One subjective measure, in the form of a questionnaire, was used to detect habituation to either of the stimulus types. The questionnaire was to be filled in by the participant after the experiment and consisted of four questions. Participants were given the following definitions about S&S:

- Startle:
  - a fast response to sudden, intense stimuli
  - to cause to start involuntarily, by or as by a sudden shock
- Surprise:
  - the feeling caused by something unexpected happening
  - an emotional response to something unexpected, resulting from a mismatch between the expectations of the environment and the perception of this environment

The questions consisted of Likert scales ranging from 0 to 10. After reading the definitions, the first two questions were related to the startling stimuli. Participants had to mark a point on the scale about their subjective feeling of startle for the first and last startling stimulus that was presented; with 0 being 'not startled at all', and 10 indicating 'extremely startled'. The remaining two questions were focused on the surprising stimuli (rare non-targets). Again, participants had to mark a point indicating their experience of being surprised by the first and last surprising stimulus. 0 indicated 'not surprised at all', and 10 indicated 'extremely surprised'.

F. Dependent physiological measures

To validate physiological responses, several dependent measures were employed. Averaged EMG amplitude data post-stimulus served as the basis for validating both eye blink and increased trapezius activity. Instantaneous HR data post-stimulus was utilised to assess increased HR due to startle and the surprise HR pattern. For the surprise HR pattern, the examination focused on averaged HR within two distinct time windows: -0.2 to 1.6 seconds post-stimulus for the 1-second HR decrease, and 1.4 to 4.8 seconds post-stimulus for the subsequent HR increase, conform: [35].

Additionally, delta and theta frequency band activity were assessed using averaged EEG activity within these respective frequency bands. Two distinct methods, maximum peak-to-peak amplitude and integrated signal, were applied to calculate activity in these frequency bands. The validation of the P300 ERP involved the analysis of the averaged EEG signal and the average of the selected back-projected ICA components. The selected back-projections were compared between two independent raters using Matthew's correlation coefficient to denote inter-rater reliability on the selected components [53].

For questionnaire responses, the indicated values on Likert scales were utilised as the dependent measures. For a more comprehensive understanding of the calculation methods applied to these dependent measures, please refer to Appendix D.

G. Statistical analysis

The data of one participant were discarded due to an error in the hardware resulting in saturated values on all channels. All data of the remaining 22 participants were used for analysis. Data treatment procedures can be found in Appendix D.

For validating the occurrence of the physiological effects, paired t-tests were used if the data were normally distributed. Otherwise, Wilcoxon signed rank tests were used. The effect size was calculated using Cohen's d. Effect sizes were interpreted as follows: small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d < 0.8$) and large ($0.8 \leq d$) [54].

It is hypothesised that:

1) Increased orbicularis oculi EMG activity will be observed after a startling stimulus, compared to a baseline value before the stimulus
2) Increased trapezius EMG activity will be observed after a startling stimulus, compared to a baseline value before the stimulus
3) An increase in HR will be observed after a startling stimulus, compared to a baseline value before stimulus onset
4) After suitable noise filtering techniques, the P300 ERP will be observed after a surprising stimulus (rare non-target)
5) Increased activity in the EEG delta (1-3Hz) and theta (3-6Hz) frequency bands will be observed after a surprising stimulus, compared to a baseline value before the stimulus
6) Compared to a baseline value before stimulus, an initial decrease in HR will be observed up until around one second post-stimulus for surprising stimuli. This is followed by a subsequent increase in HR, compared to the baseline value during the decrease period, starting at around on second post-stimulus and lasting around three seconds

IV. Phase I results: validation of S&S effects

This section will deal with the results related to the validation of the occurrence of the physiological effects listed in Table I. Detailed calculation methods for all effects can be found in Appendix D.

A. Validation of startle conditions

All startle validation results are summarised in Table II.

1) Trapezius activity increase: Using a Wilcoxon signed rank test, the EMG amplitude after stimulus onset is significantly higher than before stimulus onset with a large effect size: $Z = 253.0, p < 0.01, d = 0.876$.

TABLE II: Comparison of startle effects: post-stimulus vs pre-stimulus

	Pre-stimulus Mean (SD)	Post-stimulus Mean (SD)	p	Effect size
EMG trapezius amplitude [μV²]	8.5 (41.7)	1155.0 (3578.3)	< 0.01 [i]	0.876
EMG eye amplitude [μV²]	3.9 (4.7)	75.0 (116.9)	< 0.01 [i]	0.876
Heart rate [BPM]	76.4 (10.9)	81.7 (11.3)	< 0.01 [ii]	0.523

[i] Using Wilcoxon signed rank test [ii] Using paired t-test

2) Eye blinks: A Wilcoxon signed rank test showed that the EMG activity post-stimulus was significantly greater than the pre-stimulus EMG activity with a large effect size: $Z = 253, p < 0.01, d = 0.876$.

3) Heart rate: A Shapiro-Wilk test showed that pre-stimulus data and post-stimulus data were normally distributed ($W = 0.976, p = 0.844$ and $W = 0.946, p = 0.259$). A subsequent paired t-test showed a significant increase for the post-stimulus average heart rate compared to the pre-stimulus averaged heart rate, but with a medium effect size: $t = 4.618, p < 0.01, d = 0.523$.

B. Validation of surprise conditions

1) Increased activity in EEG delta (1-3Hz) frequency band: Table III summarises the results for rare non-target stimuli, frequent non-target stimuli and frequent target stimuli respectively. INT: using the integration method for calculating activity, PTP: using the peak-to-peak method for calculating activity (see Appendix D). A significant increase in delta frequency activity was observed for rare non-target stimuli using the integration method. Both using the averaged data ($Z = 180, p = 0.04, d = 0.370$), and the data for the first stimulus only ($Z = 180, p = 0.04, d = 0.370$), both with small effect sizes.

The same results hold for frequent non-target stimuli: using the integration method, an increase in delta band activity was observed for both the averaged data ($Z = 190, p = 0.02, d = 0.440$), as for the first stimulus data ($t = 1.79, p = 0.04, d = 0.383$). Again effect sizes were small.

For frequent target stimuli, a significant increase was found only when using the integration method, with a large effect size ($t = 5.93, p < 0.01, d = 1.265$). Pre-and post stimulus data were normally distributed ($W = 0.93, p = 0.13$ and $W = 0.95, p = 0.25$ respectively).

2) Increased activity in EEG theta (3-6Hz) frequency band: Again, the results can be found in Table III. Using the integration method, a significant increase in averaged activity over stimuli was found after stimulus onset for all stimulus types using the Wilcoxon signed rank test. A large effect size occurred for rare non-target stimuli ($Z = 253, p < 0.01, d =$

TABLE III: Comparison EEG frequency band activity: post-stimulus vs pre-stimulus. INT: using integration method, PTP: using peak-to-peak method

	Rare non-target stimuli				Frequent non-target stimuli				Frequent target stimuli			
	Pre-stimulus Mean (SD)	Post-stimulus Mean (SD)	p	Effect size	Pre-stimulus Mean (SD)	Post-stimulus Mean (SD)	p	Effect size	Pre-stimulus Mean (SD)	Post-stimulus Mean (SD)	p	Effect size
Average delta activity (INT) [μV²]	2.9 (1.0)	3.2 (1.2)	0.041 [i]	0.370	3.1 (1.1)	3.4 (1.2)	0.020^{i}	0.44	2.9 (0.9)	3.4 (0.9)	< 0.01 [ii]	1.265
First stimulus delta activity (INT) [μV²]	2.6 (1.7)	3.3 (2.6)	0.041^{i}	0.370	2.5 (1.0)	3.0 (1.2)	0.044^{i}	0.383	3.0 (1.6)	3.3 (1.2)	0.153^{ii}	0.218
Average theta activity (INT) [μV²]	1.2 (0.8)	10.3 (21.9)	$< 0.01^{i}$	0.876	0.9 (0.3)	1.1 (0.3)	$< 0.01^{i}$	0.716	0.9 (0.3)	1.1 (0.3)	$< 0.01^{i}$	0.703
First stimulus theta activity (INT) [μV²]	0.9 (0.2)	7.9 (7.4)	$< 0.01^{i}$	0.876	0.9 (0.3)	1.2 (0.5)	0.00125^{i}	0.619	0.9 (0.4)	1.0 (0.5)	0.399^{i}	0.059
Average delta activity (PTP) [μV]	45.2 (21.3)	30.6 (14.9)	0.999^{i}	-0.843	53.5 (30.6)	33.7 (16.2)	1.0^{i}	-0.855	46.3 (18.8)	31.7 (11.3)	1.0^{i}	-1.471
First stimulus delta activity (PTP) [μV]	41.3 (33.6)	32.0 (30.2)	0.980^{i}	-0.440	39.6 (21.9)	27.1 (14.3)	0.995^{i}	-0.550	52.4 (33.6)	28.6 (13.1)	0.999^{i}	-0.682
Average theta activity (PTP) [μV]	57.0 (86.6)	252.3 (588.8)	$< 0.01^{i}$	0.876	34.9 (16.8)	21.4 (8.1)	0.999^{i}	-0.536	33.6 (14.5)	20.1 (6.7)	1.0^{i}	-0.876
First stimulus theta activity (PTP) [μV]	33.9 (18.4)	176.7 (164.5)	$< 0.01^{i}$	0.855	29.2 (17.3)	21.6 (10.5)	0.995^{i}	-0.536	39.0 (26.4)	17.6 (9.2)	0.999^{i}	-0.772

[i] Using Wilcoxon signed rank test [ii] Using paired t-test

TABLE IV: Surprise HR pattern analysis

	HR difference mean (SD)	p	Effect size
ΔHR for intial drop a w.r.t. stimulus type			
Rare non-target [BPM]	-0.18 (2.29)	0.847^{i}	-0.218
Frequent non-target [BPM]	-0.45 (0.82)	0.990^{ii}	-0.534
Frequent target [BPM]	0.17 (0.67)	0.137^{ii}	0.240
ΔHR for secondary increase w.r.t. stimulus type			
Rare non-target [BPM]	0.09 (2.35)	0.434^{ii}	0.036
Frequent non-target [BPM]	1.26 (1.10)	$< 0.01^{ii}$	1.117
Frequent target [BPM]	2.19 (1.38)	$< 0.01^{ii}$	1.551

[i] Using Wilcoxon signed rank test [ii] Using single sample t-test

0.876), while medium effects sizes were observed for frequent non-target ($Z = 230, p < 0.01, d = 0.716$) and frequent target ($Z = 228, p < 0.01, d = 0.703$) stimuli.

Considering responses to the first stimulus of each stimulus type, a significant increase in integrated theta activity was found for rare non-target stimuli and frequent non-target stimuli; with a large effect size for rare non-target stimuli ($Z = 253, p < 0.01, d = 0.0.876$) and a medium effect size for frequent non-target stimuli ($Z = 216, p < 0.01, d = 0.619$).

Using the PTP method, a significant increase in averaged post-stimulus activity, averaged over all stimuli, was found compared to pre-stimulus activity for rare non-target stimuli only. A large effect size ($Z = 253, p < 0.01, d = 0.876$) was observed.

Considering responses to the first stimulus of each stimulus type, a significant increase in PTP theta activity post-stimulus compared to pre-stimulus was found for rare non-targets only, with a large effect size ($Z = 250, p < 0.01, d = 0.855$).

3) Heart rate pattern: Results regarding the HR pattern analysis can be found in Table IV. Considering a decrease in HR, no significant effect was found for either stimulus type. Secondary increase data were normally distributed for rare non-targets ($W = 0.98, p = 0.84$), frequent non-targets ($W = 0.93, p = 0.15$) and frequent targets ($W = 0.93, p = 0.10$). A significant increase in HR was found for frequent non-target ($t = 5.24, p < 0.01, d = 1.117$) and frequent target stimuli ($t = 7.27, p < 0.01, d = 1.551$), both having large effect sizes.

4) P300 ERP: The averaged EEG response, averaged over all stimuli, participants and electrodes (Fz, Cz, Pz) for a given stimulus type can be seen in Figure 4. Coloured fields indicate

the standard deviation. RNT: rare non-target, FNT: frequent non-target, FT: frequent target. No clear peak at 300 ms can be observed for the rare non-target stimuli. In addition, no clear peak suggesting a P300 ERP can be observed for the other stimulus types.

The averaged EEG response to the first stimulus of each type, averaged over participants and electrodes, can be seen in Figure 5. No clear peak at 300 ms representing the P300 can be seen for either stimulus type.

Having no clear peak at around 300 ms indicates that the P300 can either not be retrieved by averaging over three electrodes, or that the P300 wave simply does not occur after rare non-target stimuli.

In both figures, the signal amplitude ranges at around $10\mu V$, which corresponds to the ERP signal amplitude, see Appendix D4. However, this is no proof that the averaged signal indeed results in the ERP signal.

Regarding the variance in Figures 4 and 5, variance is high for all signals. Lower variance at $t = 0s$ can be explained by a pre-stimulus baseline being substracted from the signal, causing all signals to start at a relatively equal value. Increasing variance can be attributed to the noisy EEG signal and possible electrode drift. A use case of independent component analysis (ICA) is to eliminate this variance by extracting underlying components in the noisy signal [55].

Considering the back-projected independent components, 38 selections were made on rare non-target stimuli, 25 for frequent target stimuli and 25 for frequent target stimuli (see Appendix D for the selection procedure). The averaged time traces of these selected back-projections, averaged over electrodes and participants, can be seen in Figure 6. A clear peak occurring at 300 ms can be seen for all stimulus types, signalling the P300 ERP for all stimulus types. No clear distinction between stimulus types can be made using ICA with subsequent back-projection. However, the averaged waveform for the rare non-target stimuli is sharper, denoting the need for a stringent selection on the back-projected signals. The averaged signal amplitudes lie in the $10\mu V$ range, corresponding well to the ERP signal amplitude and not differing greatly from the amplitudes in Figures 4 and 5. In addition, variance is lower, indicating that the noise components were removed by the ICA decomposition. A higher variance value for the rare non-targets may be attributed to the selection criteria allowing components with an amplitude of a few dozen microvolts (see Appendix D4). The Matthew's correlation coefficient between two independent raters was 0.02, indicating an almost completely random agreement. The percentage of agreement was 50.1%, denoting that selection of components can be ambiguous and multiple waveforms might possibly correspond tot the P300, depending on the rater.

C. Post-experiment questionnaire

Table V summarises the questionnaire results. A significant decrease in the subjective ratings of both startle and surprise

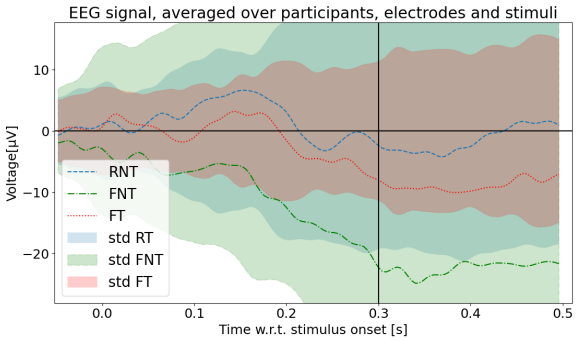


Fig. 4: Averaged EEG signal over stimuli for every stimulus type

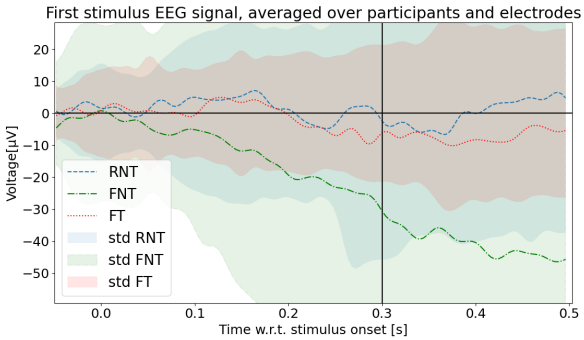


Fig. 5: Averaged EEG response to the first stimulus

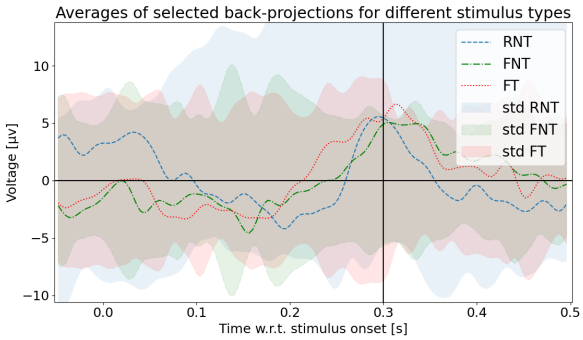


Fig. 6: Averages of selected back-projections per stimulus type

TABLE V: Post-experiment questionnaire results

	First stimulus mean (SD)	Last stimulus mean (SD)	p	Effect size
Startle	8.548 (1.120)	5.339 (1.811)	< 0.01[i]	-1.657
Surprise	6.643 (1.795)	1.187 (1.700)	< 0.01[ii]	-0.877

[i] Using paired t-test [ii] Using Wilcoxon signed rank test

is found for the last presented stimulus, compared to the first presented stimulus. Both effects sizes are large.

D. Phase I summary

The first phase of this report served to validate the startle and surprise conditions and their physiological effects. This was achieved using an experiment to induce S&S and measure the difference in physiological effects listed in Table I. All hypothesised physiological effects related to startle could be observed during the experiment.

Regarding surprise, The P300 ERP was not clearly extracted by averaging out the raw EEG leads. Using ICA, the P300 wave could be made visible in the back-projected traces. It was shown to occur for all stimulus types, but was selected more often for rare non-target responses. A clearer P300 waveform was visible for rare non-target stimuli, indicating the need for a stringent selection. Increased delta frequency band activity was observed for all stimulus types using the integration method. Theta band activity increased for all stimulus types using integration to denote activity, with results only differing in effect size. A larger effect was observed for rare non-target stimuli. When using the maximum peak-to-peak amplitude, increased theta band activity could only be observed to rare non-target stimuli. The specific HR pattern to be elicited by surprising stimuli could not be reliably reproduced. Instead, an increase in HR with a large effect size was found for frequent target and non-target stimuli, instead of for rare non-target stimuli.

V. Phase II method: detection algorithm

The detection algorithm served to answer the following research question:

1) Is it possible to detect startle & surprise by creating a detection algorithm based on the physiological effects listed in Table I?

Instead of relying on machine learning or artificial intelligence techniques, which typically demand more data than available in this experiment, we applied appropriate filtering and threshold methods to pre- and post-stimulus data for each effect individually.

A. Algorithm construction

The detection algorithm was constructed in such a way that it uses the results provided by the validation of physiological effects in Section IV. The detection method for each significant physiological effect will be described in this section. If no filtering method is described for an effect, the filtering and preprocessing was equal to the methods described in Appendix D. All data was segmented around stimulus onset. Table VII shows the pre-and post-stimulus duration of the segments per data type. Determination of threshold parameters will be discussed in Section VI.

1) Startle detection:

a) EMG orbicularis oculi activity (eye blinks) : The average amplitude of 1 s pre-stimulus ($A_{pre,emgE}$) was compared against the average of the 0-200 ms post-stimulus amplitude ($A_{post,emgE}$). If the post-stimulus amplitude exceeded the pre-stimulus amplitude by a certain factor (T_{eye}), such that $A_{post,emgE} > A_{pre,emgE} \cdot T_{emgE}$, a detection was counted.

b) Increased EMG trapezius ativity: Detection for increased trapezius activity is similar to the detection of eye blinks using EMG. The average amplitude of 1 s pre-stimulus ($A_{pre,emgT}$) was compared against the average of the 50-400 ms post-stimulus amplitude ($A_{post,emgT}$). If the post-stimulus amplitude exceeded the pre-stimulus amplitude by a certain factor ($T_{trapezius}$), such that $A_{post,emgT} > A_{pre,emgT} \cdot T_{trapezius}$, a detection was counted.

c) Eye blinks using EEG: In addition to eye blink detection using EMG, eye blink detection was also performed using the EEG Fp1 channel amplitude. Raw EEG Fp1 data was first band-pass filtered between 3 and 25 Hz to remove possible ECG artefacts. Afterwards, the amplitude of the signal was computed in the same way as for the EMG channels, described in Appendix D. By doing this, the detection of eye blinks using EEG happened in a similar fashion to the EMG detections. The average amplitude of 2 s pre-stimulus ($A_{pre,eegFp1}$) was compared against the average amplitude of 0-200 ms post-stimulus ($A_{post,eegFp1}$). If the post-stimulus amplitude exceeded the pre-stimulus amplitude by a certain factor (T_{eye_2}), such that $A_{post,eegFp1} > A_{pre,eegFp1} \cdot T_{eye_2}$, a detection was counted.

d) Increased HR: To detect increased HR, the instantaneous HR computed per 200 ms intervals (as explained in Appendix D) was compared between pre-and post stimulus using a threshold method. First, the quality of the ECG signal was checked for all segments using the *ecg_eventrelated()* Neurokit function. If data quality was higher than 0.7, segments were considered for detections. The average HR of 5 s of pre-stimulus data in BPM ($A_{pre,HR}$) was compared against the average HR in the 3-5 s post-stimulus interval ($A_{post,HR}$). If the post-stimulus HR was larger than the pre-stimulus average by a certain BPM threshold ($T_{HR,startle}$) , such that $A_{post,HR} > A_{pre,HR} + T_{HR,startle}$, a detection was counted.

e) Detection: If two or more effects were separately detected for a given stimulus, the stimulus was considered as a startling stimulus.

2) Surprise detection:

a) Increased EEG frequency band activity: Since both the delta and theta band showed increased activity for surprising stimuli (although not exclusive to surprising stimuli, see Section IV-B), surprise detection was based on a combination of increased delta and theta frequency band activity.

Tables III shows a higher increase in delta band activity for frequent non-target and frequent target stimuli compared to rare non-target stimuli. Thus and upper threshold is used for the increase in delta frequency band activity. For each segment, the delta band activity was compared against the pre-stimulus activity using only the integration method. A 2 s baseline

was used and compared against the 0-500 ms post-stimulus activity. If the integrated signal post-stimulus ($I_{post,delta}$) was higher than the integrated signal pre-stimulus ($I_{pre,delta}$) by a certain factor (T_{delta}) such that: $\frac{I_{pre,delta}}{4} < I_{post,delta} < \frac{I_{pre,delta}}{4} \cdot T_{delta}$, a detection for increased integrated activity was flagged. Division by a factor 4 was performed to re-scale the integrated signal from 2 s to 500 ms.

For theta band activity, the activity post stimulus was compared against the pre-stimulus activity using both the integration method and the max PTP amplitude method. This was again done for every segment. A 2 s baseline was used and compared against the 250-500 ms post-stimulus activity. If the integrated signal post-stimulus ($I_{post,theta}$) was higher than the integrated signal pre-stimulus ($I_{pre,theta}$) by a certain factor (T_{theta_I}) such that: $I_{post,theta} > \frac{I_{pre,theta}}{8} \cdot T_{theta_I}$, a detection for increased integrated activity was flagged. Division by a factor 8 was performed to re-scale the integrated signal from 2 s to 250 ms.

Using the PTP method, detection occurred in a similar way. If the maximum PTP amplitude in a 2 s pre-stimulus window ($A_{pre,theta}$) was smaller than the maximum PTP amplitude in a 250-500 ms post-stimulus window ($A_{post,theta}$) by a certain factor (T_{theta_P}) such that if: $A_{post,theta} > A_{pre,theta} \cdot T_{theta_P}$, a detection for increased PTP activity was flagged. A detection of increase theta activity was flagged if both the integrated and PTP activity detections were flagged. Delta activity was treated separately.

b) HR decrease and increase: Since the results of Table IV do not comply with [35], the surprise HR detection was based on the only significant effect from Table IV: if the average HR in the 1.4 - 4.8 s interval post-stimulus is higher than the average HR in the -0.2-1.6 s interval post-stimulus, we do not detect surprise. This is implemented by setting a threshold indicating the maximum increase in average HR in the 1.4-4.8 s interval. For every segment, the averaged HR in the -0.2-1.6 s interval was calculated (HR_{base}). If the average HR in the 1.4-4.8 s interval (HR_{incr}) is not higher than HR_{base} by a certain BPM value ($T_{HR,surprise}$), such that if: $HR_{incr} < T_{HR,surprise} + HR_{base}$, a surprise detection was counted.

c) P300 ERP: Detection of the P300 ERP was performed in a similar manner to Appendix D, but with all 78 stimulus responses considered instead of the 6 frequent targets, 6 frequent non-targets and 6 rare non-targets. No averaging of the raw EEG traces was used, selection was performed using ICA and back-projection. Again, selection of the independent components which might represent the P300 ERP was performed by the researcher. A subsequent selection on the back-projections was performed. Selected back-projections were counted as surprise detections. It is important to note that selection of components and back-projections happened in a blind manner: the researcher did not know the type of stimulus when performing the selection for a certain segment.

d) Detection: If two or more separate surprise effects were detected for a given stimulus, the stimulus was considered a surprise stimulus.

TABLE VI: Threshold parameters and their value after grid search

Parameter	Considered range [start, stop, step]	Selected value
Startle		
T_{eye}	[1, 7, 0.2]	3.6
$T_{trapezius}$	[0.2, 5, 0.2]	2.6
T_{eye_2}	[2, 8, 0.2]	5.0
$T_{HR,startle}$	[1, 10.8, 0.2]	11.8
Surprise		
T_{delta}	[1025, 2, 0.025]	1.625
T_{theta_I}	[1, 2, 0.025]	1.7
T_{theta_P}	[1, 2, 0.025]	1.05
$T_{HR,surprise}$	[-5, 5, 0.2]	4

TABLE VII: Duration of segments pre-and post-stimulus

	Pre-stimulus [s]	Pre-stimulus [s]
ECG	3	6
EMG amplitude	1	1
EEG Fz, Pz, Cz (used in ICA)	1	1
EEG Fz, Pz, CZ (used for frequency bands)	2	2
EEg Fp1 amplitude (eye blink detection)	2	2

3) Parameter search: A parameter search was performed on the threshold parameters from Table VI using a grid search. The considered ranges for every parameter are provided in Table VI. Correct and false detections were computed for very parameter value considered. Parameter values were chosen based on the following cost function:

$$Q(x, i) = x \cdot C[i] - (1 - x) \cdot F[i], \tag{1}$$

where Q is a $1 \times n$ row vector with n the number of considered parameter values for a certain parameter. The i^{th} element represents the cost for choosing the i^{th} parameter value from its considered range. C is a $1 \times n$ row vector containing the average number of correct detections, averaged over all participants and stimuli, corresponding to every possible parameter value. The i^{th} element thus represents the average number of correct detections when using the i^{th} parameter value. Similarly, F is a $1 \times n$ row vector containing the average number of false detections corresponding to every possible parameter value. x is a scalar value representing the balance given to correct or false detections. For a given value of x, the ideal parameter value for a certain parameter from Table VI can then be found by maximising $Q(i)$ for i.

By varying x, the focus of the algorithm can be changed. Higher x values focus more on classifying as many stimuli as possible, ignoring false detections. Lower values of x have the adverse effect: fewer false detections but a more stringent classification. An x value of 0.76 was chosen for the startle parameter search, and an x value of 0.94 for the surprise parameter search. These values were chosen because they give an average correct classification of at least 5 startle, respectively, 1.5 surprise stimuli when performing detection using the verification data set.

TABLE VIII: Average correct and false startle detections out of 6 startle stimuli

Effect (method)	Average correct detections (SD)	Average false positive detections (SD)
Verification data set, N = 72		
Trapezius activity (EMG)	4.9 (1.2)	0.42 (0.64)
Eye blink (EMG)	4.3 (1.7)	1.3 (1.2)
Eye blink (EEG)	4.8 (1.7)	1.6 (2.1)
Increaseed HR (ECG)	1.6 (1.8)	5.0 (3.8)
Startle	5.0 (1.2)	0.83 (0.99)
Validation data set, N = 60		
Trapezius activity (EMG)	4.7 (1.9)	0.6 (1.2)
Eye blink (EMG)	4.9 (1.6)	1.7 (1.3)
Eye blink (EEG)	5.0 (1.9)	0.7 (1.0)
Increaseed HR (ECG)	1.1 (1.4)	4.8 (4.6)
Startle	5.1 (1.9)	0.20 (0.40)

B. Data usage

To ensure that the algorithm does not overfit the obtained data, a portion of the data was used to perform the parameter search and extract optimal parameters, while the remaining data was used to validate the detection algorithm. Data from 12 participants was used for the parameter search, from now on denoted as the verification set. Data from the other 10 participants were used for validation, from now on denoted as the validation set.

Six startle stimuli, six rare non-target stimuli and 33 frequent target and frequent non-target stimuli were presented per participant. This resulted in a total of 72 startle stimuli, 72 rare non-target stimuli, and 396 frequent target and frequent non-target stimuli for the verification set. Similarly, 60 startle stimuli, 60 rare non-target stimuli, and 330 frequent target and frequent non-target stimuli were used for the validation data set.

VI. Phase II results: detection algorithm

A. Verification results

The obtained parameter values using the verification data set can be found in Table VI.

1) Startle detections: Table VIII summarises the startle detection results, using the parameters of Table VI. Results denote the number of correct and false startle detections per participant, averaged over all participants. Averaged over all participants, a detection rate of 5.0 stimuli (SD = 1.2 stimuli) of the 6 presented startling stimuli was achieved, with an average false detection rate of 0.83 stimuli (SD = 0.99 stimuli).

2) Surprise detections: Table IX summarises the surprise detection results, using the parameters of Table VI. Averaged over all participants, a detection rate of 3.6 stimuli (SD = 0.9 stimuli) of the 6 presented surprising (rare non-target) stimuli was achieved, with an average false detection rate of 39.8 stimuli (SD = 7.8 stimuli). The first rare non-target stimulus had an average detection rate of 0.6 (SD = 0.5), indicating that on average, the first rare non-target stimulus was detected 60% of the time. False detections are very high, with more than half of the presented stimuli falsely classified as surprising.

TABLE IX: Average correct and false surprise detections out of 6 surprise stimuli

Effect (method)	Average correct detections (SD)	Average false positive detections (SD)
Verification data set, N = 72		
Increased delta band activity (EEG)	1.3 (1.1)	18.1 (8.7)
Increased theta band activity (EEG)	3.0 (0.9)	25.2 (4.2)
P300 ERP (EEG)	1.2 (1.0)	10.7 (4.2)
HR pattern (ECG)	5.2 (0.8)	63.8 (5.5)
Surprise	3.6 (0.9)	39.8 (7.8)
First surprise stimulus	0.6 (0.5)	
Validation data set, N = 60		
Increased theta band activity (EEG)	2.1 (0.9)	23.9 (6.7)
Increased delta band activity (EEG)	1.9 (0.8)	29.1 (5.0)
P300 ERP (EEG)	1.3 (0.9)	7.9 (2.1)
HR pattern (ECG)	5.0 (0.9)	64.0 (4.8)
Surprise	3.7 (1.2)	42.9 (5.4)
First surprise stimulus	0.8 (0.4)	

B. Validation results

1) Startle detections: Table VIII summarises the validation startle detection results, using the parameters of Table VI. Averaged over all participants, a detection rate of 5.1 stimuli (SD = 1.9 stimuli) of the 6 presented startling stimuli was achieved, with an average false detection rate of 0.2 stimuli (SD = 0.4 stimuli). Figure 7 shows the percentage of correct detections per stimulus number. Detection rates are high for all stimuli, but a 10% drop can be seen after the third stimulus. Table X shows the confusion matrix for startle detection using the validation set. In total 78 stimuli were presented per participant, of which 6 were startling stimuli. Results are expressed in average detected stimuli, averaged across participants False positive and false negative detections are low (0.2 and 0.9 stimuli per participant respectively).

	Detection	No detection
Startle stimulus	5.1	0.9
No startle stimulus	0.2	71.8

TABLE X: Confusion matrix for startle detection validation

2) Surprise detections: Table IX summarises the surprise detection results, using the parameters of Table VI. Averaged over all validation participants, a detection rate of 3.7 stimuli (SD = 1.2 stimuli) of the 6 presented surprising (rare non-target) stimuli was achieved, with an average false detection rate of 42.9 stimuli (SD = 5.4 stimuli). The first rare non-target stimulus had an average detection rate of 0.8 (SD = 0.4). Again, false positive detections are very high. Figure 8 shows the percentage of correct detections per stimulus number. There exists no apparent difference for later stimuli in terms of detection rate. Table XI shows the surprise detection confusion matrix for the validation set. Of the 78 presented stimuli per participant, 6 were rare non-targets (surprising stimuli). Results are expressed in average detected stimuli, averaged across participants. False positive detections are high,

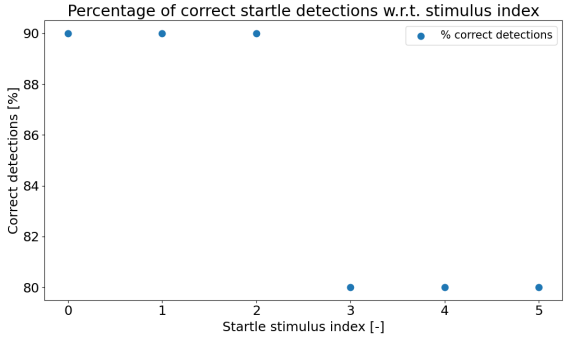


Fig. 7: Percentage of correct startle detections w.r.t. stimulus number

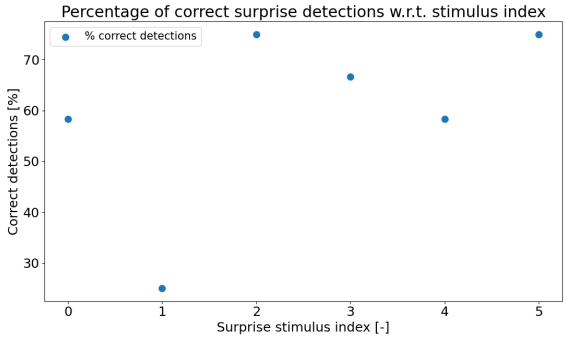


Fig. 8: Percentage of correct surprise detections w.r.t. stimulus numbe

with on average 42.9 out of 72 false stimuli falsely classified as surprising.

	Detection	No detection
Startle stimulus	3.7	2.3
No startle stimulus	42.9	29.1

TABLE XI: Confusion matrix for surprise detection validation

VII. Discussion

This section will separately discuss the findings regarding validation of the physiological effects, and the performance of the detection algorithm. After, a general discussion placed in the wider research context, and directions for further research are presented.

A. Questionnaire responses

The responses to the S&S questionnaire showed that the used stimulus trains indeed elicited S&S in participants. This validates the use of the stimulus trains at hand. It should be noted that the subjective feeling of surprise was lower than that of startle, indicating that participants were less surprised than they were startled.

B. Validation of physiological effects.

1) Startle: Hypotheses 1), 2) and 3) from Section III-G can thus be confirmed. Since both trapezius and orbicularis oculi activity increased with a large effect size, these effects can be reliably produced by the used startle stimuli and can be accurately measured. Increased HR after stimulus onset had a significant but smaller effect size. This effect can thus also be reliably induced by the used startle stimuli, but to a lesser extent.

2) Surprise: Since surprise is a more cognitive process than startle, greater difficulty was expected in finding reliable indicators. Hypothesis 4) from Section III-G is confirmed, with the caveat that the P300 ERP is not exclusive to surprising stimuli. It is thus no reliable indicator for surprise. This is reflected in the P300 ERP not being clearly extracted by averaging out the raw EEG leads, even if the first, and expected most surprising, rare non-target stimulus response was used for averaging. Using ICA whereby independent components could be selected and inverted for back-projection, the P300 wave could be made visible in the back-projected traces. However, the P300 was shown to occur for all stimulus types. The P300 was selected more often for rare non-target responses (30.8%, versus 19.8% for frequent non-targets and 20.3% for frequent targets). A clearer P300 waveform was visible for rare non-target stimuli, indicating the need for a stringent selection. The P300 ERP proved not to be solely elicited by surprising stimuli. In addition, the inter rater reliability proved to be almost completely random, signalling that selecting components is ambiguous. A suitable explanation might be that the P300 ERP is not elicited by surprise nor novelty, but rather by stimuli that require mental effort to be processed. Mixed results can be found in literature to support this claim [56], [57]. Cognitive effort might also be needed when frequent target stimuli occur, since the participant needs to update their mental count. Due to the specific instructions given to ignore the frequent non-target stimuli, participants might still consciously process these stimuli, again requiring cognitive effort. Another explanation would be that not enough ICA components could be extracted to effectively have one component representing the ERP. This may result in multiple unrelated signals combining into one component [58], which then looks like the P300 ERP, giving a false image. A solution to this would be to use more EEG electrodes, since the number of independent components is limited by the number of sensors [55]. Due to the limited number of electrodes, ICA can not always find a global solution, which can result in inversion of back-projected components [58].

Considering the hypothesised increased activity in the EEG delta (1-3 Hz) and theta (3-6 Hz) frequency bands, mixed results were found. Hypothesis 5) from Section III-G can be confirmed for the theta band, but once again with the caveat that the increase is not necessarily exclusive for rare non-target stimuli. The hypothesis is rejected for the delta band activity. Both delta and theta frequency band activity did increase, depending on its calculation method. An increased

delta frequency band activity was observed for all stimulus types using the integration method, and for no stimuli by using the PTP method. Effect sizes were small for both rare non-target and frequent non-target stimuli, but large for frequent target stimuli. This makes the use of the integrated delta band activity two-fold since it is not exclusive to surprise stimuli. In addition, the effect is not greater for rare non-target stimuli, thus no distinction can be made based on effect size. Delta band activity is thus no indicator of surprise.

Theta band activity increased for all stimulus types using integration to denote activity, with results only differing in effect size. A larger effect was observed for rare non-target stimuli. When using the maximum peak-to-peak amplitude, increased theta band activity could only be observed to rare non-target stimuli. This makes increase theta band activity a good secondary indicator for surprise. Responses to the first rare non-target stimulus did not provide different results, indicating that the effect of the first stimulus might not be as pronounced as first expected. A possible explanation might be that the P300 ERP contains many frequencies in the 3-6 Hz range. Since the P300 ERP is shown to not only elicited by surprising stimuli, increases in theta band activity will follow this pattern.

Hypotheses 6) from Section III-G can not be confirmed. The specific HR pattern to be elicited by surprising stimuli, whereby a HR decrease lasting around 1 s is followed by an increased HR lasting about 3 s, could not be reliably reproduced. This is in sharp contrast to [35]. Instead, an increase in HR with a large effect size was found for frequent target and non-target stimuli, instead of for rare non-target stimuli. The specific HR pattern is thus no indicator of surprise.

C. Detection algorithm performance

1) Startle: Due to the physiological effects related to startle being clearly present, high detection performance was expected. This is reflected in Table VIII. Similar values for the verification and validation detection rate indicate that there is no overitting on the verification data set. The algorithm shows a reliable detection of most startle stimuli, with a low number of false detections (Table X). This balance between false and correct detections can also further be shifted by shifting the weight parameter x in the cost function of Eqn. 1, making the algorithm flexible in use.

Since participants indicated being less startled after the last stimulus compared to the first (see Table V), a logical drop in detection performance for later stimuli could be expected. This is reflected in Figure 7, which shows a drop in detection performance of 10% starting from the third stimulus. These results are also in line with Table V, indicating a lesser experience of startle for later stimuli. Nonetheless, the reflexive nature of startle is clear due to the high detection rate and thus physiological response for every stimulus.

False detections may be attributed to multiple causes: individual differences, involuntary or unconscious usage of muscles, and anticipatory startle. Individual differences between participants can not be avoided, as each individual will react slightly different to (the absence of) stimuli. Since detection was partly based on muscle usage (orbicularis oculi, trapezius), participant movement and unconscious eye blinks can be a source of false EMG and EEG Fp1 detections. Anticipatory startle can occur, where the effects already occur in the build-up to a stimulus, since participants expect to be startled but don't know when [5], [49]. This can lead to false detections as well.

2) Surprise: Due to surprise being less well described by the physiological effects listed in Table I, lower detection performance was expected. This is in line with the obtained results from Table IX. No clear differences between the verification and validation data set performance was observed, indicating that no overfitting on the verification data occurred and the algorithm generalises well. Numerical results will be discussed for the validation data set. On average, only 3.7 of 6 stimuli (61.7%) could be detected, whereas on average more than 42 of the remaining 72 stimuli were falsely detected a surprising stimuli. The number of false positive detections was high (Table XI), indicating that the effects also occurred without surprising stimuli.

A possible explanation for the poor performance is the fact that participants generally reported lower levels of surprise than startle, indicating that their physiological reactions might be less pronounced. In addition, not all effects listed in Table I were proven to occur after surprising events, making detection inherently harder.

Participants indicated being less surprised to later stimuli (see Table V), thus a higher detection rate for the first stimulus was expected. This also did not prove to be the case, since no clear trend in degrading performance for later stimuli can be observed (Figure 8). On average, 60% of the first stimuli were correctly classified, but false positives were very high.

D. General discussion

The implemented detection algorithm can successfully detect startle to a high accuracy, with only one set of parameters fit for all participants. This indicates that the startle detections can be used as an objective indicator of startle to be used alongside subjective ratings obtained by the S&S scale currently being developed. It can assist in developing the scale by giving a more comprehensive overview of whether startle occurs, and provide which effects are detected. By having a properly developed subjective scale, novel pilot training scenario's including startle can be verified and validated for effectiveness in inducing startle. This may then lead to pilots being more capable of dealing with startling situations on the flight deck.

Surprise was not detected accurately by the algorithm. The current implementation can not be used to provide a ground truth for developing the S&S scale. More research is needed to successfully elicit and detect surprise using physiological data, which will be discussed next.

E. Recommendations for future work

Eye tracking can be incorporated into the sensor suite. This eye tracker can be used to detect startle eye blinks without the

need for EMG/EEG measurements. Since eye tracking can be invasive, a desk-mounted tracker would be preferable. Eye tracking also enables the measurement of gaze, which is linked to both surprise [8], [33] and the orienting startle response [5], [18].

To reliably detect the P300 ERP, a first step could be to use more electrodes. Since having more electrodes allows for extracting more independent components using ICA [55]. This increases chances of effective noise reduction and having one independent component describing purely the ERP [58], [59]. In addition, averaging over more electrodes reduces the influence of noise, thereby making the averaged raw EEG signal more useful. Again, invasiveness should be kept to a minimum.

An alternative way of detecting surprise would be to first detect which physiological parameters change when participants are presented with a surprising stimulus. This can be achieved using anomaly detecting on the whole data set. To achieve this, more data should be collected since these methods generally use machine learning.

To accurately detect surprise, a separate experiment should be used without the startling stimuli present. This eliminates the possible influence of participant being distracted or mentally consumed by the startling stimuli and can thus reduce possible confounds which were inevitable due to the time constraints of this project.

VIII. Conclusion

An experiment designed to induce startle and surprise in participants was performed to validate the physiological effects induced by S&S and to construct an automatic detection algorithm. All physiological effects related to startle were reliably reproduced and automatic startle detection could be performed with high accuracy. Regarding surprise, only some effects were observed. These effects are: the P300 ERP, and an increase in EEG theta frequency band (3-6 Hz) activity. Due to not all effects being observed and not being exclusive to surprising stimuli, detection accuracy for surprise was significantly lower, and more research needs to be performed to find reliable indicators which can be measured in a non-obtrusive manner. The findings in this report can help develop the subjective S&S scale by providing a ground truth for startle.

Appendix

A. Stimulus train generation

Stimulus trains were constructed in a pseudo-random way, with care being taken that startling stimuli did not occur within three seconds prior to or after a visual stimulus. In addition, the first ten visual stimuli were forced to be frequent stimuli, not rare non-targets stimuli. Three stimulus trains were constructed for startle, and four stimulus trains for surprise.

B. Detailed apparatus information

The board contains 8 channels to be used for either EMG, ECG or EEG, and has a sampling rate of 250 Hz [60]. The board is provided with a Bluetooth-to-serial converter, which was plugged into a Dell Precision 5820 workstation to ensure communication between the board and PC.

EEG electrodes on the Fz, Cz and Pz locations were dry EEG reusable comb electrodes, with Ten20 conductive gel applied. The EMG electrodes on the orbicularis oculi and the EEG Fp1 electrode were Kendall 24 mm Ag/AgCI disposable electrodes. For trapezius EMG and ECG, Kendall 35 mm Ag/AgCI disposable electrodes were used. The OpenBCI GUI was used to check proper connection of all electrodes and to verify the impedance of the EEG electrodes before the start of the experiment [61].

C. EEG electrode impedances

Table XII provides the mean impedance per EEG electrode and its standard deviation across participants.

TABLE XII: Impedance of EEG electrodes

	Average [$k\Omega$]	Standard deviation [$k\Omega$]
Fz	375.3	236.7
Cz	385.0	257.4
Pz	385.4	442.9
Fp1	35.2	22.3

D. Data treatment for validation of physiological effects

1) EMG activity: EMG data were first filtered using a 50 Hz notch filter to remove electrical interference. Afterwards, a 60 Hz high-pass Butterworth filter of order 5 was used with a forward and backward pass to remove time delays. After, the signal was de-trended in the 0^{th} and 1^{st} order to remove bias and drift. The amplitude of the EMG signal was then calculated using the linear envelope of the signal. The EMG amplitude, averaged over all 6 startling stimuli, before and after stimulus onset were compared for all participants ($N = 22$) for both the orbicularis oculi and trapezius channels. For EMG trapezius analysis, the average of 1 s of pre-stimulus data was used as a pre-stimulus baseline. Post-stimulus activity was averaged in the 50-400 ms post-stimulus interval. Regarding the EMG orbicularis oculi channel, the averaged signal of 1 s pre-stimulus was used as a baseline, and the average amplitude from 0-200 ms post-stimulus was used as post-stimulus activity. The two sets of 22 averaged values each were subsequently compared against each other using a Wilcoxon signed rank test or a paired t-test, depending on the normality of both sets.

2) Heart rate: The ECG data were first filtered using 50 Hz and 60 Hz notch filters to remove electrical interference. After, a 0.67 - 30 Hz band-pass Butterworth filter of order 5 was used with a forward and backward pass. Following this, the signal was de-trended in the 0^{th} and 1^{st} order to remove bias and drift. The Neurokit API [62] was then used to further clean the ECG signal using the *ecg_clean()* function.

HR was calculated by first extracting the R-peaks using the Neurokit *ecg_process()* function. Using the R-peaks, instantaneous real-time heart rate was computed for 200 ms intervals as in [35] using the calculation method from [63].

For detecting increased HR due to startle, 5 s pre-stimulus and 5 s post-stimulus heart rate was averaged over all startling stimuli ($N = 6$) for every participant. For both methods, pre- and post-stimulus data were averaged over all surprise stimuli ($N = 6$) for all participants ($N = 22$). The two sets of 22 averaged values each were subsequently compared against each other using a Wilcoxon signed rank test or a paired t-test, depending on the normality of both sets.

For detecting the surprise HR pattern, consisting of a 1 s decrease and followed by a 3 s increase, the 3 s before stimulus onset were averaged and used as baseline for detecting a decrease in HR. The averaged HR from -0.2-1.6 s post-stimulus was subtracted from this baseline to detect a decrease in HR. If the difference proved to be significantly greater than 0, we confirm a decrease in the -0.2-1.6 s interval. The -0.2-1.6 s post-stimulus interval served as the baseline for detecting the increase in HR. The averaged HR in the -0.2-1.6 s post stimulus window was subtracted from the average HR in the 1.4-4.8 s post-stimulus window to detect an increased HR. If the difference proved to be significantly greater than 0, an increased HR in the 1.4-4.8 s post-stimulus window was confirmed. This method is similar to [35], the exact same intervals are used here. Data were averaged over all rare non-target stimuli ($N = 6$) for all participants ($N = 22$). In addition, data corresponding to the first surprise stimulus were separately analysed. For each condition (increase or decrease), the two set of 22 values each were subsequently compared against each other using a Wilcoxon signed rank test or a paired t-test, depending on the normality of both sets.

3) EEG frequency band activity: Fz, Cz and Pz EEG data were used to detect the increase in delta and theta frequency band activity. The Fp1 channel was not used due to it containing possible HR and eye blink artefacts. For delta frequency band activity, the raw data were filtered using a 1-3 Hz band-pass Butterworth filter of order 4 with a forward and backward pass. For the theta band, filtering consisted of a 3-6 Hz band-pass Butterworth filter of order 5 with a forward and backward pass.

Two methods for calculating the frequency band activity were used: integration and maximum peak-to-peak (PTP) amplitude.

The first consisted of a digital integration of the signal using the composite trapezoidal rule with a spacing between sample points of 0.004. A 2 s pre-stimulus baseline was integrated and compared against the integrated signal of the 0-500 ms post-stimulus signal for the delta band; and the 250-500 ms post-stimulus signal for the theta band.

Using the PTP method, the maximum PTP amplitude in a 2 second pre-stimulus window was compared against the maximum PTP amplitude in a 0-500 ms post-stimulus window for the delta band; and a 250-500 ms window for the theta band.

For both methods, pre-and post-stimulus data were averaged over all surprise stimuli ($N = 6$) for all participants ($N = 22$). The two sets of 22 averaged values each were subsequently compared against each other using a Wilcoxon signed rank test or a paired t-test, depending on the normality of both sets.

4) P300 ERP: The Fz, Cz and Pz channels were used to detect the P300 ERP. the raw data were first filtered using 50 Hz and 60 Hz band stop filters to remove electrical interference. After, the signal was de-trended in the 0^{th} and 1^{st} order. This was followed by a 0.1 - 35 Hz band-pass Butterworth filter with order 5 which was used in a forward and backward pass. Segments containing 1 s of pre-stimulus and 1 s of post-stimulus data were used. For validating the P300 response, two methods were used: averaging the filtered EEG signal, and using independent component analysis (ICA) with subsequent back-projection [58], [59].

Averaging the filtered EEG data started with subtracting a 100 ms pre-stimulus baseline from the signal. The remaining signal up until 800 ms post-stimulus was then averaged across electrodes, stimuli and participants.

Using ICA, the same 100 ms baseline was subtracted from the signal. A random selection of six frequent non-target responses, six frequent target responses and six rare non-target responses were used. Responses in which eye blinks occurred were discarded. Eye blinks were detected using the EEG method described in Section V. On average, the responses to 5.6 rare non-target, 5.7 frequent non-target and 5.6 frequent target stimuli were used. For each response, FastICA was performed on the signal using three independent components. This was done to separate the background noise from the ERP signal. The components were shown in an interactive graphical user interface where the researcher was able to select components possibly corresponding to the P300 ERP (if any) for each response. In addition, the averaged EEG trace over electrodes, and the column averages of the mixing matrix A [55] were shown, indicating a polarity switch when back-projecting a certain component. Selection of components was based on the following rules:

- The component should show a clear positive or negative peak between 250-400 ms post-stimulus.
- No large periodic oscillations should be visible in the 100-400 ms interval.
- The component should not have any other clear peaks in the 100-400 ms post-stimulus window, other than the peaks corresponding to the N100, N200 or P200 ERPs in both time and polarity. See Figure 9 for an example.
- If the peak round 300 ms is negative, the sign of the A matrix column average corresponding to the component should be negative as well, and vice versa.

Selected components for all responses were then back-projected to the scalp in order to re-obtain a signal in μV [58]. A second selection round was then performed according to the following rules:

- The back-projected P300 peak amplitude should not exceed a few dozen microvolts [35], [64].
- The general shape of the back-projected signal should correlate well with the P300 ERP in both time and polarity, without other peaks being shown in the 100-400 ms post-stimulus interval, except for possibly the

N100, N200 and P200 ERPs with correct time traces and amplitudes. The shape of the signal in the 100-400 ms window was compared against Figure 9.

- No clear periodic signals should be visible in the back-projected signal between 100-400 ms.

Selected back-projections were then classified as being a P300 ERP. The selected back-projections per stimulus type were averaged over all stimuli and participants. To eliminate the possibility of biased results due to the manual selection, selection was also performed by another independent researcher, which was only given the above selection criteria. The inter rater reliability was calculated using Matthew's correlation coefficient (also known as "Phi coefficient") to show the level of agreement between the two subjective selections [53].

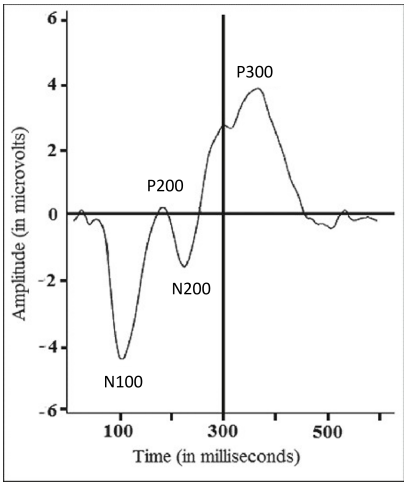


Fig. 9: Reference EEG signal for P300 selection. Image modified from [65, Figure 1]

REFERENCES

[1] Boeing Commercial Airplanes. "Statistical Summary of c Commercial Jet Airplane Cncidents: Worldwide Operations 1959 – 2020." (2020), [Online]. Available: https://skybrary.aero/sites/default/files/bookshelf/32664.pdf (visited on 12/16/2022).

[2] L. G. Mitello and R. J. B. Hutton, "Applied cognitive task analysis (acta): A practitioner's toolkit for understanding cognitive task demands," *Ergonomics*, vol. 41, no. 11, pp. 1618–1641, 1998, PMID: 9819578.

[3] M. S. Young and N. A. Stanton, "Malleable attentional resources theory: A new explanation for the effects of mental underload on performance," *Human Factors*, vol. 44, no. 3, pp. 365–375, 2002, PMID: 12502155.

[4] L. Bainbridge, "Ironies of automation," in *Analysis, Design and Evaluation of Man–Machine Systems*, G. Johanssen and J. Rijnsdorp, Eds., Pergamon, 1983, pp. 129–135.

[5] M. Koch, "The neurobiology of startle," *Progress in Neurobiology*, vol. 59, no. 2, pp. 107–128, 1999.

[6] C. Belcastro and J. Foster, "Aircraft loss-of-control accident analysis," in *AIAA Guidance, Navigation, and Control Conference*. 2010.

[7] J. Burki-Cohen, "Technical challenges of upset recovery training: Simulating the element of surprise," in *AIAA Modeling and Simulation Technologies Conference*. 2010.

[8] R. Reisenzein, G. Horstmann, and A. Schützwohl, "The cognitive-evolutionary model of surprise: A review of the evidence," *Topics in Cognitive Science*, vol. 11, no. 1, pp. 50–74, 2019.

[9] K. Vicente and J. Rasmussen, "Ecological interface design: Theoretical foundations," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 4, pp. 589–606, 1992.

[10] J. Kuchar and L. Yang, "A review of conflict detection and resolution modeling methods," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 4, pp. 179–189, 2000.

[11] European Aviation Safety Agency. "Notice of Proposed Amendment 2015-13: Loss of control prevention and recovery training." (2015), [Online]. Available: https://www.easa.europa.eu/en/document-library/notices-of-proposed-amendment/npa-2015-13 (visited on 12/16/2022).

[12] Federal Aviation Administration. "AC 120-111 - Upset Prevention and Recovery Training - with Change." (2015), [Online]. Available: https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/1027328 (visited on 12/16/2022).

[13] International Civil Aviation Organisation, *Manual of evidence-based training (doc 9995)*, English, version 1, ICAO, 2013, 162 pp.

[14] A. Landman, E. L. Groen, M. M. (van Paassen, A. W. Bronkhorst, and M. Mulder, "Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise," *Human Factors*, vol. 59, no. 8, pp. 1161–1172, 2017, PMID: 28777917.

[15] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, ser. Advances in Psychology, P. A. Hancock and N. Meshkati, Eds., vol. 52, North-Holland, 1988, pp. 139–183.

[16] F. Dehais, M. Causse, and J. Pastor, "Embedded eye tracker in a real aircraft: New perspectives on pilot/aircraft interaction monitoring," 2008.

[17] J. Rivera, A. B. Talone, C. T. Boesser, F. Jentsch, and M. Yeh, "Startle and surprise on the flight deck: Similarities, differences, and prevalence," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, no. 1, pp. 1047–1051, 2014.

[18] W. L. Martin, P. S. Murray, P. R. Bates, and P. S. Y. Lee, "A flight simulator study of the impairment effects of startle on pilots during unexpected critical events," *Aviation Psychology and Applied Human Factors*, vol. 6, no. 1, pp. 24–32, 2016.

[19] E. Siedlecka and T. Denson, "Experimental methods for inducing basic emotions: A qualitative review," *Emotion Review*, vol. 11, p. 175 407 391 774 901, Mar. 2018.

[20] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological Psychology*, vol. 84, no. 3, pp. 394–421, 2010, The biopsychology of emotion: Current theoretical and empirical perspectives.

[21] M. K. Noordewier, S. Topolinski, and E. Van Dijk, "The temporal dynamics of surprise," *Social and Personality Psychology Compass*, vol. 10, no. 3, pp. 136–149, 2016.

[22] Y. E. Dreissen, M. J. Bakker, J. H. Koelman, and M. A. Tijssen, "Exaggerated startle reactions," *Clinical Neurophysiology*, vol. 123, no. 1, pp. 34–44, 2012.

[23] C. Ladd, P. Plotsky, and M. Davis, "Startle response," *Encyclopedia of stress*, vol. 3, pp. 561–568, Jan. 2010.

[24] D. E. Wilkins, M. Hallet, and M. M. Wess, "Audiogenic Startle Reflex Of Man And Its Relationship To Startle Syndromes: A Review," *Brain*, vol. 109, no. 3, pp. 561–573, Jun. 1986.

[25] S. Holand, A. Girard, D. Laude, C. Meyer-Bisch, and J.-L. Elghozi, "Holand s, girard a, laude d, meyer-bisch c, elghozi jl. effects of an auditory startle stimulus on blood pressure and heart rate in humans. j hypertens 17: 1893-1897," *Journal of hypertension*, vol. 17, pp. 1893–7, Jan. 2000.

[26] M. Kuhn, J. Wendt, R. Sjouwerman, C. Büchel, A. Hamm, and T. B. Lonsdorf, "The neurofunctional basis of affective startle modulation in humans: Evidence from combined facial electromyography and functional magnetic resonance imaging," *Biological Psychiatry*, vol. 87, no. 6, pp. 548–558, 2020, Stress Mechanisms and Fear Memories.

[27] C. C. Duncan-Johnson and E. Donchin, "On quantifying surprise: The variation of event-related potentials with subjective probability," *Psychophysiology*, vol. 14, no. 5, pp. 456–467, 1977.

[28] C. Başar-Eroglu, T. Demiralp, M. Schürmann, and E. Başar, "Topological distribution of oddball 'p300' responses," *International Journal of Psychophysiology*, vol. 39, no. 2, pp. 213–220, 2001.

[29] C. Başar-Eroglu, E. Başar, T. Demiralp, and M. Schürmann, "P300-response: Possible psychophysiological correlates in delta and theta frequency channels. a review," *International Journal of Psychophysiology*, vol. 13, no. 2, pp. 161–179, 1992.

[30] J. Harper, S. M. Malone, and W. G. Iacono, "Theta- and delta-band eeg network dynamics during a novelty oddball task," *Psychophysiology*, vol. 54, no. 11, pp. 1590–1605, 2017.

[31] C. Mulert, O. Pogarell, G. Juckel, *et al.*, "The neural basis of the p300 potential: Focus on the time-course of the underlying cortical generators," *European archives of psychiatry and clinical neuroscience*, vol. 254, pp. 190–8, Jul. 2004.

[32] K. Preuschoff, B. 't Hart, and W. Einhauser, "Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making," *Frontiers in Neuroscience*, vol. 5, 2011.

[33] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision Research*, vol. 49, no. 10, pp. 1295–1306, 2009, Visual Attention: Psychophysics, electrophysiology and neuroimaging.

[34] R. Reisenzein, S. Bördgen, T. Holtbernd, and D. Matz, "Evidence for strong dissociation between emotion and facial displays: The case of surprise," *Journal of personality and social psychology*, vol. 91, pp. 295–315, Sep. 2006.

[35] R. F. Simons, F. K. Graham, M. A. Miles, and M. T. Balaban, "Input and central processing expressed in erp and heart rate changes to rare target and rare nontarget stimuli," *Psychophysiology*, vol. 35, no. 5, pp. 563–575, 1998.

[36] F. K. Graham, "Attention: The heartbeat, the blink, and the brain," in *Attention and information Processing in infants and Adults*, Psychology Press, 1992, pp. 3–29.

[37] J. Blascovich and W. Mendes, "Feeling and thinking: The role of affect in social cognition," *Studies in Emotion and Social Interaction*, pp. 59–82, Jan. 2000.

[38] J. Healey and R. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, 2005.

[39] J. Tichon, T. Mavin, G. Wallis, T. Visser, and S. Riek, "Using pupillometry and electromyography to track positive and negative affect during flight simulation," *Aviation Psychology and Applied Human Factors*, vol. 4, p. 23, Jan. 2014.

[40] T. D. Blumenthal, B. N. Cuthbert, D. L. Filion, S. Hackley, O. V. Lipp, and A. Van Boxtel, "Committee report: Guidelines for human startle eyeblink electromyographic studies," *Psychophysiology*, vol. 42, no. 1, pp. 1–15, 2005.

[41] K. Ryu and R. Myung, "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic," *International Journal of Industrial Ergonomics*, vol. 35, no. 11, pp. 991–1009, 2005.

[42] F. Rooseleer, B. Kirwan, E. Humm, and D. Alarcon, "'the application of human factors in wake vortex encounter flight simulations for the reduction of flight upset risk and startle response," Jan. 2022.

[43] Z. Gibson, J. Butterfield, M. Rodger, B. Murphy, and A. Marzano, "Use of dry electrode electroencephalography (eeg) to monitor pilot workload and distraction based on p300 responses to an auditory oddball task," in *Advances in Neuroergonomics and Cognitive Engineering*, H. Ayaz and L. Mazur, Eds., Cham: Springer International Publishing, 2019, pp. 14–26.

[44] M. K. Johnson, J. A. Blanco, R. J. Gentili, K. J. Jaquess, H. Oh, and B. D. Hatfield, "Probe-independent eeg assessment of mental workload in pilots," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2015, pp. 581–584.

[45] O. Bruna, T. Levora, and J. Holub, "Assessment of ecg and respiration recordings from simulated emergency landings of ultra light aircraft," *Scientific Reports*, vol. 8, May 2018.

[46] M. Hannula, K. Huttunen, J. Koskelo, T. Laitinen, and T. Leino, "Comparison between artificial neural network and multilinear regression models in an evaluation of cognitive workload in a flight simulator," *Computers in Biology and Medicine*, vol. 38, no. 11, pp. 1163–1170, 2008.

[47] J. A. Veltman, "A comparative study of psychophysiological reactions during simulator and real flight," *The International Journal of Aviation Psychology*, vol. 12, no. 1, pp. 33–48, 2002.

[48] OpenBCI. "OpenBCI Cyton Biosensing Board (8-channels)." (2023), [Online]. Available: https : / / shop . openbci . com / products / cyton - biosensing-board-8-channel.

[49] *Handbook of Psychophysiology* (Cambridge Handbooks in Psychology), 4th ed. Cambridge University Press, 2016.

[50] G. Rojas, C. Alvarez, C. Montoya Moya, M. de la Iglesia Vaya, J. Cisternas, and M. Gálvez, "Study of resting-state functional connectivity networks using eeg electrodes position as seed," *Frontiers in Neuroscience*, vol. 12, Mar. 2018.

[51] C. C. Duncan, R. J. Barry, J. F. Connolly, *et al.*, "Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, p300, and n400," *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1883–1908, 2009.

[52] S. Zavala, K. Chicaiza, J. Murillo López, J. Ramírez, and S. G. Yoo, "Eeg signal processing model for eye blink detection," *Journal of Engineering and Applied Sciences*, vol. 15, pp. 1671–1675, Mar. 2020.

[53] M. J. Grant, C. M. Button, and B. Snook, "An evaluation of interrater reliability measures on binary tasks using d-prime," *Applied Psychological Measurement*, vol. 41, no. 4, pp. 264–276, 2017.

[54] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas," *Frontiers in Psychology*, vol. 4, 2013.

[55] G. Naik and D. Kumar, "An overview of independent component analysis and its applications," *Informatica*, vol. 35, pp. 63–81, Jan. 2011.

[56] F. Howells, D. Stein, and V. Russell, "Perceived mental effort correlates with changes in tonic arousal during attentional tasks," *Behavioral and brain functions : BBF*, vol. 6, p. 39, Jul. 2010.

[57] A.-M. M. P. Ullsperger and H.-G. Gille, "The p300 component of the event-related brain potential and mental effort," *Ergonomics*, vol. 31, no. 8, pp. 1127–1137, 1988, PMID: 3191898.

[58] F. Cong, I. Kalyakin, and T. Ristaniemi, "Can back-projection fully resolve polarity indeterminacy of independent component analysis in study of event-related potential?" *Biomedical Signal Processing and Control*, vol. 6, no. 4, pp. 422–426, 2011.

[59] C. Bugli and P. Lambert, "Comparison between principal component analysis and independent component analysis in electroencephalograms modelling," *Biometrical Journal*, vol. 49, no. 2, pp. 312–327, 2007.

[60] OpenBCI. "OpenBCI Documentation: Cyton Board." (2022), [Online]. Available: https://docs.openbci.com/Cyton/CytonLanding/.

[61] OpenBCI. "Installing the OpenBCI GUI as a "Standalone" Application." (2023), [Online]. Available: https://docs.openbci.com/Software/OpenBCISoftware / GUIDocs / #installing - the - openbci - gui - as - a - standalone-application.

[62] D. Makowski, T. Pham, Z. J. Lau, *et al.*, "NeuroKit2: A python toolbox for neurophysiological signal processing," *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, Feb. 2021.

[63] F. K. Graham, "Constraints on measuring heart rate and period sequentially through real and cardiac time," *Psychophysiology*, vol. 15, no. 5, pp. 492–495, 1978.

[64] R. B. Mars, S. Debener, T. E. Gladwin, *et al.*, "Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise," *Journal of Neuroscience*, vol. 28, no. 47, pp. 12 539–12 545, 2008.

[65] S. Samima, M. Sarma, and D. Samanta, "Detecting vigilance in people performing continual monitoring task," Dec. 2017, pp. 202–214.

Part II

Preliminary Graduation Report

*This part has been assessed for the course AE4020 Literature Study.

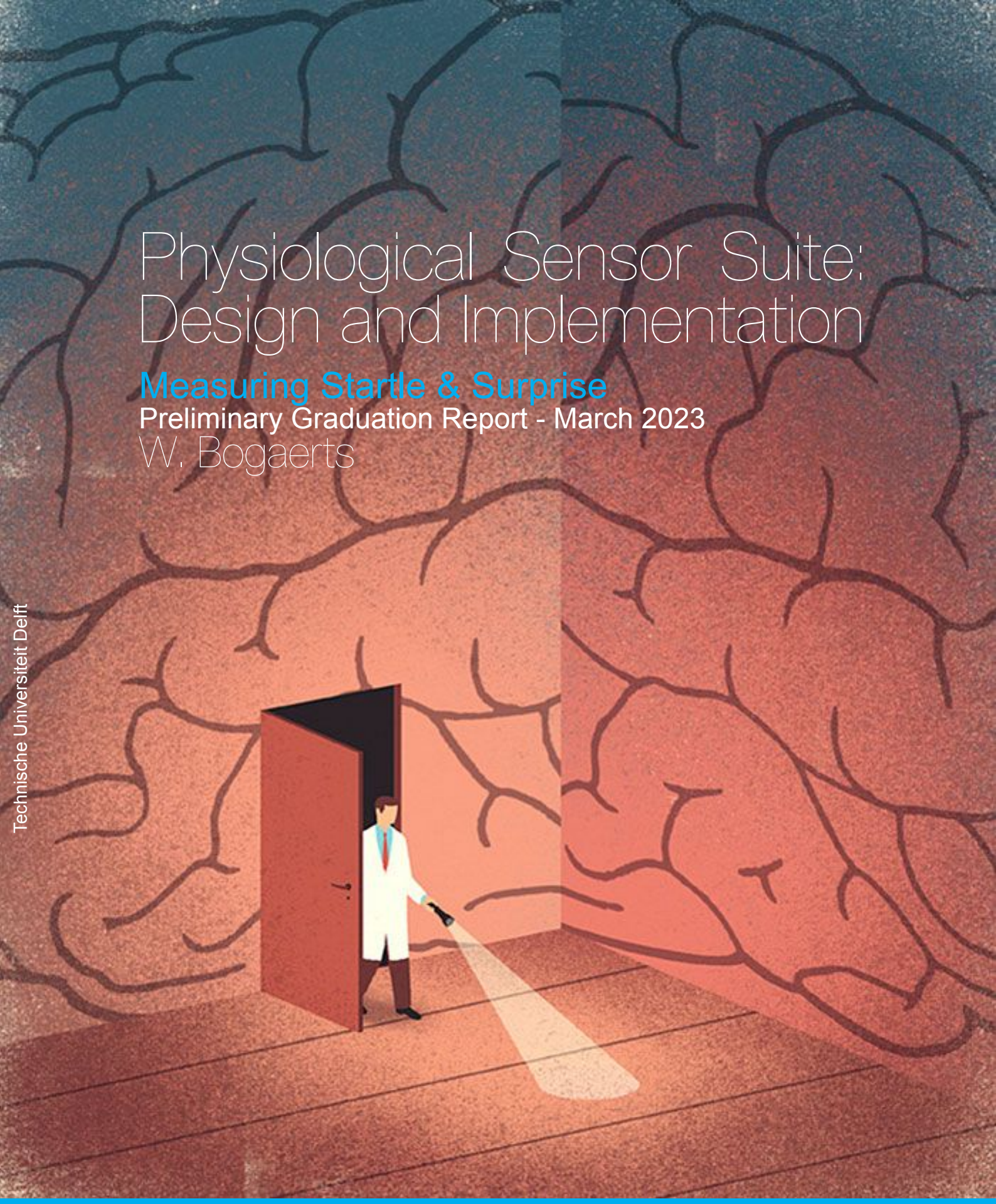
Physiological Sensor Suite: Design and Implementation

Measuring Startle & Surprise

Preliminary Graduation Report - March 2023

W. Bogaerts

Technische Universiteit Delft



TU Delft

Cover image by Davide Bonazzi, https://www.davidebonazzi.com/

Physiological Sensor Suite: Design and Implementation

Measuring Startle & Surprise

by

W. Bogaerts

Preliminary Master Thesis of Science
Delft University of Technology,
March 10, 2023

Student number:	5378273	
Readers:	Prof. dr. ir. M. Mulder,	TU Delft
	Dr. ir M. M. van Paassen,	TU Delft, supervisor
	Ir. O. Stroosma,	TU Delft
	Dr. A. H. Landman,	TNO, TU Delft

TU Delft

List of Figures

List of Tables

List of abbreviations

ACC anterior cingulate cortex. 8, 10

ADC analog-to-digital converter. 5, 6, 22–25

API application programming interface. 17–23, 27, 28

BLE Bluetooth low energy. 17, 19–21, 23, 26, 27, 30

BP blood pressure. 6–8, 10, 13

BVP blood volume pulse. 6

CMRR common mode rejection ratio. 22–25

DUECA Delft University Environment for Communication and Activation. 17, 18, 20, 23, 26–28, 30, 31, 33

ECG electrocardiography. iii, v, ix, 7, 11–13, 15, 16, 18–20, 24–26, 30, 31, 33

EDA electrodermal activity. 5, 6, 13, 27

EEG electroencephalography. iii, v, ix, 8–13, 15, 16, 18–21, 23, 24, 26, 29–33

EMG electromyography. iii, v, ix, 5, 10–13, 15, 16, 18–23, 25–27, 29–31, 33

ERP event-related potential. 8, 9, 11, 13, 15, 18, 31–33

HR heart rate. 7, 8, 10, 11, 13, 15, 16, 29–31, 33

HRV heart rate variation. 7, 11, 13, 15, 16, 29–31, 33

ICA independent component analysis. 9, 11, 12

PNS parasympathetic nervous system. 7, 10, 16

PPG photoplethysmography. 6, 7, 13

RMS root mean square. 21, 22, 24, 25, 30

S&S startle & surprise. 1, 3, 11, 15, 16, 29, 31–33

SC skin conductance. iii, 5, 6

SCL skin conductance level. iii, 5, 6

SCR skin conductance response. iii, 4–6, 13

SNR signal-to-noise ratio. 22, 27

SNS sympathetic nervous system. 5, 7, 10, 15, 16

SRS SIMONA research simulator. 8, 11, 15, 17, 26

Contents

1 Introduction

With increased levels of automation autonomy and authority in the flight deck, the role of the pilot has become more supervisory (Wiener & Curry, 1980). However, in situations that can not be handled by automated systems, the role of the pilot becomes critical and manual control may be needed. These situations tend to be unforeseen and complex, inducing the need for rapid adaptation to the situation and quick judgement and decision making (Mitello & Hutton, 1998). The switch to an active role can be difficult after e.g. long periods of automated flight (Young & Stanton, 2002). Furthermore, pilots' manual flying skills degrade with extensive use of automation. In addition, over-reliance on the automated system can decrease the level of active monitoring by the pilot. Due to the automation not being completely transparent, pilots can not obtain full knowledge of the process executed by the automation which can lead to automation surprises. (Bainbridge, 1983)

Loss of control still forms the largest category of fatal aviation accidents (Boeing Commercial Airplanes, 2020), and due to the pilot being kept out of the loop, a startle factor can arise during unexpected situations. This startling effect complicates the crew's ability to troubleshoot the problem (Belcastro & Foster, 2010; Burki-Cohen, 2010). To reduce the issue that startling and surprising effects impose, two main directions have been explored by previous research. The first deals with preventing the problem by increasing the quality of the automated systems. An example of research in this direction is the development of ecological interfaces to give the pilot insight into the physics behind the automation, thereby increasing the automation transparency (Borst et al., 2006; Vicente & Rasmussen, 1992). Other examples include already implemented automated safety systems such as conflict detection and resolution systems (Kuchar & Yang, 2000). If, however, surprising or startling situations still occur, the pilot profits from having the appropriate skills necessary to deal with the situation. This is the main motive for the second research direction: improving pilot training to include startle and surprise (European Aviation Safety Agency, 2015; Federal Aviation Administration, 2015; International Civil Aviation Organisation, 2013).

Understanding the influence of startle & surprise (S&S) allows to construct more effective training programs, thereby reducing the risk of fatal accidents caused by unexpected events. These events induce both cognitive and physical reactions (Rivera et al., 2014). A conceptual model of startle and surprise has already been proposed by Landman et al to explain pilot performance in surprising and startling situations (Landman et al., 2017a). In addition, a subjective rating scale to indicate startle and surprise, similar to the NASA TLX scale for measuring workload (Hart & Staveland, 1988), is being developed. A set of physiological measurements can assist in developing this scale by introducing objective quantitative data in addition to the subjective data.

The main goal of this work is to develop and implement a physiological sensor suite in the SIMONA flight simulator to measure the physiological effects of startle and surprise and assist in developing the subjective scale. This is achieved by first discovering the physiological indicators of startle and surprise. For every indicator, feasibility of implementing a sensor for this indicator inside the simulator is determined. After, the sensors are integrated into a sensor suite and implemented in both the hard-and software of SIMONA.

The academic relevance of this work is reflected in the fact that future research will be able to use an extra set of measurements to develop the startle and surprise scale. This scale can then be used to further investigate the effects of startle and surprise. Understanding these effects can lead to the construction of more effective training programs, resulting in pilots having less difficulty in dealing with surprising situations and thereby increasing safety.

The remainder is structured as follows: first, a literature review is given in Section 2. the review deals with the physiological and cognitive effects of startle & surprise and how to measure them. Section 3 proposes the research questions and hypotheses. In Section 4, the design requirements and theoretical development of the sensor suite is discussed. Lastly, Section 6 provides the expected results and overall conclusions.

2

Literature review

This chapter provides a background on S&S in the field of aviation. With this theoretical background, the physiological and cognitive effects arising from S&S are discussed. Since this work focuses on measuring physiological effects, a review on how to measure these effects is provided as well. Finally, ways of implementing these measurements are discussed.

2.1. Startle & surprise in aviation

Startle is defined as a fast response to sudden, intense stimuli, normally introduced to protect the organism from harm (Koch, 1999). It generates fast, automatic and highly physiological reaction called the startle reflex, combined with a slower conditioned and behavioural reaction called the startle response (Martin et al., 2016; Rivera et al., 2014). Attention is shifted towards the source of the startling stimulus (Rivera et al., 2014). Surprise is defined as an emotional response to something unexpected, resulting from a mismatch between the expectations of the environment and the perception of this environment. It forces reevaluation of the situation and changing a person's understanding. Surprise causes the interruption of e person's current task and induces an orienting attention shift to the cause of the surprise (Landman et al., 2017a; Rivera et al., 2014). While startle is always a response to a certain stimulus, surprise can also be the consequence of the absence of a stimulus (Rivera et al., 2014). An example would be an odd event such as a dial indicating a unexpected value. A conceptual model of S&S was developed by Landman et al. (2017a), which formulates that one's perception and actions are sometimes guided by learned frames (also called schema's). Performance issues can occur when someone is insufficiently able to adapt to a new frame. The model entails the effects of both startle & surprise and uses it to explain the response to unexpected events in the cockpit.

Startle & surprise is able to degrade pilot performance significantly. Startle is thought to introduce a considerable delay in the decision making Martin et al. (2016) and can have disorienting effects (Rivera et al., 2014). In addition, the motor performance is affected as well (Rivera et al., 2014). Surprise has the effect of increasing the difficulty of managing upset situations during flight (Landman et al., 2017b; Rivera et al., 2014).

Because of the safety concerns about S&S, future pilot training will have to include these effects (European Aviation Safety Agency, 2015; Federal Aviation Administration, 2015; International Civil Aviation Organisation, 2013). To construct these training methods as effectively as possible, some research has been performed in constructing training scenarios including startle & surprise. Landman et al. (2018) has proven that organising training scenarios in a more unpredictable way improves the transfer of training to unexpected situations in flight. Furthermore, having a mnemonic-type procedure for managing startling and surprising situa-

tions proved to increase the level of decision-making, albeit that initial responses proved to be less optimal (Landman et al., 2020).

2.2. Cognitive effects of startle & surprise

Surprise has extensive cognitive effects, such as the inability to analyse or comprehend the situation, freezing and losing situational awareness (Rivera et al., 2014). The cognitive effects of startle are an impairment of the information processing, negative impact on decision making and brief disorientation (Rivera et al., 2014). A feeling of fear or anger can be induced, together with an increased level of arousal and arrest of ongoing behaviours (Landman et al., 2017a). In addition, attentional mechanisms are focused towards the startling source (Martin et al., 2016; Sänger et al., 2014).

Both startle and surprise can result in an acute increase of stress (Martin et al., 2016). Stress is known to impair the function of the goal-directed attentional system. It increases the extent to which processing is performed by a stimulus-driven attentional mechanism, resulting in a more bottom-up approach of analysing the situation. The behaviour of a person becomes more reactive than proactive and response to distractions is increased (Eysenck et al., 2007; Landman, 2019; stephanie E. Wemm, 2017). This results in problems regarding goal-oriented planning, problem analysis and focusing on task-relevant stimuli (Landman, 2019). The stress response has a higher cognitive impact in situations with high mental load. Anxiety can further reduce mental processing (Eysenck et al., 2007).

2.3. Physiological effects of startle & surprise

Both startle and surprise will be treated separately. For each of the physiological effects, the constancy and sensitivity of this effect will be reviewed. Constancy indicates that a person is able to control the level of the physiological effect under normal circumstances. Sensitivity indicates that the value of this physiological response is indeed related to startle or surprise. A good physiological indicator fulfils both the constancy and sensitivity criteria: it can be controlled by a participant at rest but changes after a surprising or startling event/stimulus.

2.3.1. Startle

The startle response originates in the caudal brain stem and is composed of two separate responses: the initial motor response (also called muscular tension reflex or startle reflex), and the orienting response (also called startle response) (Dreissen et al., 2012). The initial motor response is fast and has an onset latency of 20-50ms. It is characterised by muscle contraction in a fight-or-flight reaction to the stimulus. No control over this response can be achieved, it is purely reflexive. The response starts with eye blinking, and is followed by a forward head movement, facial grimacing, shoulder elevation, arm abduction, elbow bending, forearm pronation, finger flexion, abdominal contraction and knee flexion (Ladd et al., 2010; Rivera et al., 2014). Most muscle activity is located in the face and shoulders, notably the blinking reflex and tension of the trapezius muscles (Dreissen et al., 2012; Landman et al., 2017a; Martin et al., 2016). The orienting response is used to focus attention on the startling stimulus and has a onset latency of 400-450ms. It is determined by the nature of the stimulus, causing a source of variance. Contrary to the initial motor response, the orienting response is an emotional and voluntary response (Dreissen et al., 2012). It involves orienting to the stimulus source, increased skin conductance (SCR), increased systolic blood pressure, heart rate and pupil size (Dreissen et al., 2012; Ladd et al., 2010; Martin et al., 2016). The initial motor response is valence-related: increased during negative emotional states and decreased during positive emotional states (Holand et al., 2000). On the contrary, the orienting response is not valence-related (Kuhn et al., 2020). In the following, the constancy and sensitivity of

the physiological effects will be discussed.

The eye blink reflex is sensitive in nature. It is a pure reflex to the startling stimulus and is directly correlated. In terms of constancy, a small caveat exists. Since people need to blink regularly, they will occur in the absence of startling stimuli as well. However, the blinking reflex is of a larger magnitude than regular blinking (Leuchs et al., 2019). Still, the resulting blinking data should be accompanied by the time stamp of the startling stimulus to clearly link the blink response to the stimulus. The eye blink response can be measured either by using electromyography (EMG) of by using eye tracking (Blumenthal et al., 2005; Leuchs et al., 2019; Tichon et al., 2014). Measuring the eye blink response using EMG has been documented by Blumenthal et al. (2005). Key requirements for good data acquisition are high enough sampling rate (>1000Hz), an amplifier with high input impedance ($>100M\Omega$), high common-mode rejection ratio (>100dB) and low input noise ($<1V$ RMS in the 10-500Hz range). A high resolution analog-to-digital converter (ADC) (>16 bits) is a plus.

Regarding shoulder muscle activity, the same principle holds. Trapezius muscles are especially sensitive to the sympathetic nervous system (SNS) response (Wijsman et al., 2010), but can also be used by the participant during e.g. control tasks. In controlled settings where participants are in a relaxed position, constancy is guaranteed. However, when participants are moving, care should be taken to include the time step of the stimulus when looking at muscle activity output. The trapezius muscle is a good indicator of startle in relaxed and controlled conditions (Dreissen et al., 2012), but should be combined with other measures in other conditions. The same principle holds for other muscles mentioned above. Depending on the experiment, appropriate muscles which are expected to not be continuously used during the experiment should be chosen to perform measurements on. Measuring muscle activity is performed using EMG. EMG is also used for measuring general muscle activity, so the technical requirements are similar to those mentioned for measuring eye blinks. Since eye blinking is a very fast response, the measurement speed is of highest importance. This results in needing a sufficient sampling rate in order to detect these fast changes.

Orientation towards the stimulus can be ignored by the participant. However, this is mostly the case when the participant has habituated to the stimulus. It is thus generally sensitive to startle. Orientation of attention can be fully controlled by a participant and thus fulfils the constancy requirements. Measurement of this index is however more difficult. The activation of muscles that are used to position the head can be measured, or a face tracking system can be used. In scenarios where the stimulus is inside the visual field of the participant, eye tracking can be used as well. In this case, a sampling rate of at least 50Hz is recommended to be able to track eye movement towards the stimulus (Carter & Luke, 2020).

Skin conductance (SC) (also called electrodermal activity (EDA)) is directly related to the sympathetic nervous system (SNS) activity. It consists of a slow changing skin conductance level (SCL) and a fast changing skin conductance response (SCR) component. The phasic SCR component is a reaction to novel, unexpected, significant or aversive stimuli (Handbook of Psychophysiology, 2016). In addition, SCR's can also occur in absence of stimuli, such as when taking a deep breath. The analysis of SCR is thus twofold. It is the only physiological reaction that only corresponds to the discrete activation of the SNS, but is not only evoked by startling stimuli. This makes SCR useful when other parameters are controlled, and only a startling stimulus is provided. Under resting conditions, generally no SCR spikes occur. Constancy and sensitivity are guaranteed if conditions are controlled well. If not, the time stamp of the startling stimulus needs to be provided in order to couple the SCR spike to the startle response. The tonic component, SCL is related to a more general state of arousal and is generally kept constant by a participant. It reacts more to chronic stimuli, such as performing a task, high workload and anger. SCL level is less useful in the quantification of startle. It

is however always recorded, so separating SCR from SCL in the skin conductance data is necessary. Large individual differences in SC exist, so categorising responses in a percentage of the individual's SC range is commonly used (Handbook of Psychophysiology, 2016). SCR initiates 1-4s after the stimulus has occurred and shows a peak in SC value. The peak can be described using many metrics in both time and shape. An example SC response can be seen in figure 2.1. Since the SCR is only a small portion of the SC, measurement resolution should be sufficiently high. EDA is measured by passing a small current through the skin via a pair of electrodes which are placed on the skin surface. It uses Ohm's law to measure skin conductance, the reciprocal of skin resistance, expressed in Siemens (S). A digital constant voltage system is preferred with a sampling frequency sufficiently high to capture the SCR dynamics. SCR is a fairly slow moving phenomena, with time scales of 1-4s. A sampling frequency of 2Hz or higher is thus preferred. In addition, the ADC should have a resolution sufficient to capture the small SC changes related to a SCR spike. The SCR size typically varies between $0.1 - 1\mu S$ (Handbook of Psychophysiology, 2016), so a change of $0.05\mu S$ is commonly chosen as a detection threshold (Boucsein, 2012). The total range of SC is typically between $2 - 23\mu S$ (Handbook of Psychophysiology, 2016). Depending on the total measurement range in Siemens, the resolution should be chosen such that a $0.05\mu S$ change can be detected. For example, a sensor with a $0 - 25\mu S$ range with an ADC resolution of 10 bits has a detection resolution of $\frac{25\mu S - 0S}{2^{10}} = 0.024\mu S < 0.05\mu S$, which is adequate.

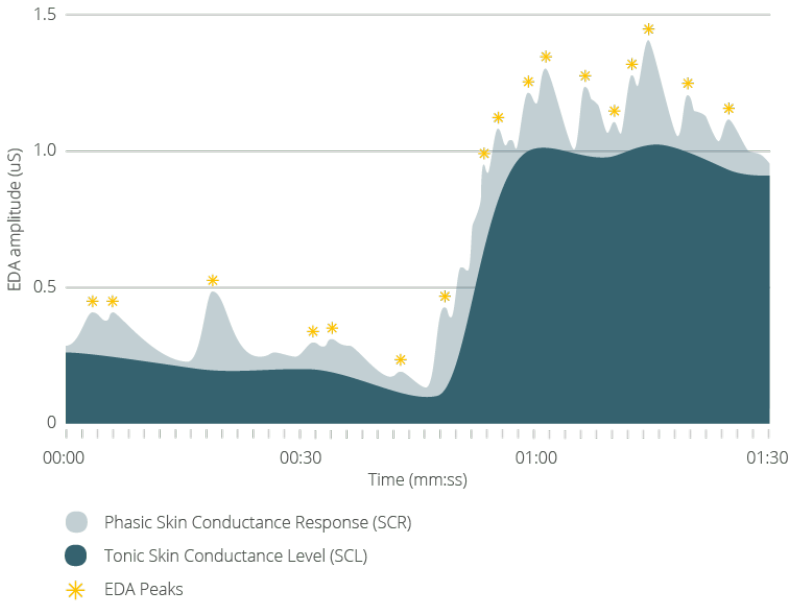


Figure 2.1: Example SC measurement with SCR peaks and SCL level indicated. Image taken from iMotions (2020)

Increased blood pressure (BP) can be caused by other factors other than startle, such as stress. It is however sensitive to startle (Holand et al., 2000) and remains constant under controlled conditions. An increase in systolic blood pressure can be observed 4-5s after the stimulus (Holand et al., 2000). BP can thus be used as a measure for startle, but not standalone just as SCR and muscle activity. BP can be measured using a variety of techniques, such as a sphygmomanometer (tonometer) (Ogedegbe & Pickering, 2010) or blood volume pulse (BVP) (also known as photoplethysmography (PPG) in this context)(Man et al., 2022). BVP is a non-invasive measurement method using a photo cell to measure blood volume in the

veins. back-scattered PPG devices are recommended, since these can be placed on the areas sensitive to SNS changes, such as the wrist (Handbook of Psychophysiology, 2016). Technical requirements are not as well defined, but Handbook of Psychophysiology (2016) recommends that any blood pressuring device must meet either the Association for the Advancement of Medical Instrumentation or British Hypertension Society standards regarding accuracy and reproducibility.

Along with BP, heart rate (HR) increases as well 3-4s after the eliciting stimulus (Handbook of Psychophysiology, 2016; Holand et al., 2000). HR is controlled by the SNS, but also by the parasympathetic nervous system (PNS), sometimes working against each other (Handbook of Psychophysiology, 2016), causing interpretation to be hard. To more accurately measure the effect of the SNS and startling stimulus, the heart rate variation (HRV) is more often used. The ratio of high frequency variability to low frequency variability provides an indication of SNS activity. Low frequency variability is driven by both the SNS and PNS, while high frequency variability is only controlled by the PNS (Handbook of Psychophysiology, 2016). This ratio thus is sensitive to the startling stimulus, but it should be compared against a reference value before the stimulus. In general, HR and HRV can be controlled by participants by e.g. breathing. The HRV and thus HR can thus be seen as a valid indicator for startle, however care should be taken. Measuring HR and HRV is commonly performed using electrocardiography (ECG). Multiple electrode configurations with up to twelve electrodes exist. However, for most psychophysiological measures, an Einthoven lead II configuration provides sufficient accuracy since they yield large R-peaks (Handbook of Psychophysiology, 2016). It uses two electrodes, placed on the right arm or chest and left leg or hip. A schematic representation of this configuration is provided in figure 2.2, along with an ECG representation of a single heartbeat. A sampling frequency of at least 500Hz is recommended to be able to detect R-peaks (Handbook of Psychophysiology, 2016; Rinold & Porges, 1997).

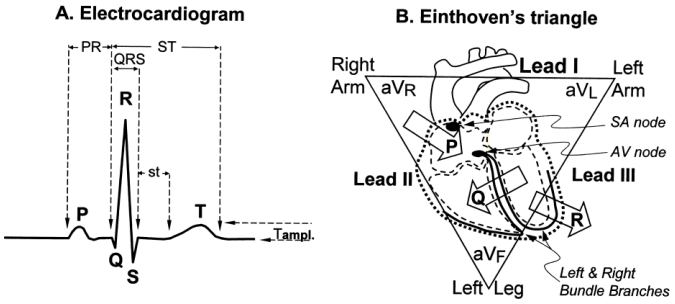


Figure 2.2: Typical ECG signal (A) three Einthoven ECG configurations (I, II, III) (B), image from (Handbook of Psychophysiology, 2016)

Pupil size has been used often to recognise the orienting startle response. It has a slow and fast changing component (Carter & Luke, 2020; Leuchs et al., 2019). The slow changing pupil size is related to general arousal and e.g. expectancy and anticipation of events. Fast pupil size increase is related to novel and arousing stimuli. Pupil dilation increases more with more arousing and novel stimuli and does not show signs of habituation (Leuchs et al., 2019). Pupil dilation due to stimuli is lower than due to sudden increased brightness (Carter & Luke, 2020). This makes that pupil dilation due to stimuli can be easily recognised. Measuring pupil dilation is often performed using eye tracking equipment. It should be noted that eye movement introduces noise in the measurements, since pupil size will adjust to different brightness in the visual field (Carter & Luke, 2020). Since pupil dilation occurs at the same time as the orientation toward the stimulus in the orienting response, eye fixation will not be guaranteed,

making it difficult to record reliable pupil dilation data. Carter and Luke (2020) suggests a fixed gaze position in order to accurately measure pupil dilation, which in the case of simulator studies in SRS can almost never be achieved. Since the pupil dilation response is fast (onset of around 200ms after the stimulus (Carter & Luke, 2020)), a sampling rate of around 1000Hz was suggested by Kret and Sjak-Shie (2018).

2.3.2. Surprise

As mentioned earlier, surprise is induced by a schema-discrepant stimulus or event (Noordewier et al., 2016; Reisenzein et al., 2019). The surprise reaction, much like the startle response, consists of two phases: the response to the unexpectedness of the event, and the sense-making of this event (Noordewier et al., 2016). Response to unexpectedness is independent of the valence of the surprising stimulus, while the sense-making can invoke different psychophysiological reaction depending on its valence (Noordewier et al., 2016). There is still a lot of discussion on the physiological effects of surprise (Kreibig, 2010; Reisenzein et al., 2019; Siedlecka & Denson, 2018). Collet et al. (1997) state that it is possible to distinguish between different emotions by using electrodermal, thermovascular and respiratory indices. This section will thus focus on reactions which have been proven to be solely elicited by surprise.

Considering the reaction to unexpectedness, an increase in the EEG P300 event-related potential (ERP) level can be observed, with larger P300 values for more schema-discrepant events (Baar-Eroglu et al., 1992; Duncan-Johnson & Donchin, 1977; Noordewier et al., 2016). The P300 response is combined with an increased activity in the delta (1-3Hz) and theta (3-6Hz) frequency bands (Baar-Eroglu et al., 2001). In addition, it increases pupil dilation (Preuschoff et al., 2011; Reisenzein et al., 2019). The P300 ERP originates in the anterior cingulate cortex (ACC) (Mulert et al., 2004), which shows an increase in activity around 300ms after the schema-discrepancy, which can be observed by a negative deflection in the EEG (Noordewier et al., 2016). It should be noted that this increased activity has also been observed in the startle response. It is therefore seen as a general indication of the aversive defense response (Hajcak & Foti, 2008). This is followed by attentional allocation and interruption of ongoing behaviours, which is characterised by action delays, increased error rate and surprise induced blindness occurring as early as 200ms after the schema-discrepancy is noticed (Noordewier et al., 2016; Reisenzein et al., 2019). Specific reactions include gaze towards the stimulus, occurring around 400ms after discrepancy (Itti & Baldi, 2009; Reisenzein et al., 2019). In addition, a surprise expression can be observed, constructed of raised eyebrows, widened eyes and open mouth (Noordewier et al., 2016). However, except for increased corrugator activity, Reisenzein et al. (2006) argues that these effects have not been reproduced sufficiently. An initial HR deceleration lasting for around 1s can be observed after the surprising stimulus (Simons et al., 1998). Between the reaction to unexpectedness and sense-making, verification of the schema discrepancy is performed by taking a second look at the surprising event (Reisenzein et al., 2019).

Sense-making is characterised by cardiovascular reactivity and depends on the nature is the event (Noordewier et al., 2016). If the event is seen as a challenge, HR will increase while BP lowers (Blascovich & Mendes, 2000). If the discrepancy is seen as a threat, increased HR and BP were observed (Blascovich & Mendes, 2000). A challenge is defined when the individual experiences (nearly) sufficient resources to deal with the situation, while a threat is characterised by an experience of insufficient resources to tackle the situation (Blascovich & Mendes, 2000). The HR increase starts at around 2s after stimulus onset (Simons et al., 1998). In a similar manner as the previous section, the constancy and sensitivity criteria of the physiological reactions will be discussed.

P300 activity is shown to be inversely proportional to the probability of an event occurring,

showing increased levels for very surprising events (Linden, 2005). In a review, Donchin et al. (2006) state that the P300 is related to the output of a mismatch detector. This confirms the sensitivity of the P300 wave to surprise. The P300 wave is only elicited on these occasions, in other controlled circumstances, P300 waves do generally not occur (Linden, 2005). The combination of these factors make the P300 ERP both subject to the constancy and sensitivity criteria. It should be noted that the P300 wave can only be observed when the stimulus time stamp is known, since it occurs around 300ms after the stimulus (Picton, 1992). This time delay can differ since the P300 is only elicited after the stimulus is categorised, for stimuli that are difficult to categorise, delays can increase to 1000ms (Duncan et al., 2009).

Duncan et al. (2009) provides extensive guidelines on measuring P300. The P300 ERP component can be measured using EEG on the Fz, Cz and Pz scalp locations with the earlobe as a reference, as shown in figure 2.3 (Duncan et al., 2009). A high-pass setting of 0.01Hz is recommended, the low-pass setting should be set to approximately 0.25 of the analog-to-digital conversion rate, with a recommended setting of 100Hz. Sampling should be performed at 200Hz minimum (Duncan et al., 2009). It should be noted that EEG, and thus the P300 component is sensitive to eye movement and blinking, especially at the frontal region of the scalp (Ghaderi et al., 2014). Filtering of the EEG signal is necessary to remove these artefacts. Using independent component analysis (ICA), eye movement artefacts can be filtered without the need for extra electrooculogram measurements (Zhou & Gotman, 2009). Although ICA can reduce the artefacts, computing has to be done after measurement data has been obtained.

Measuring the P300 response can be combined with a measurement of the delta and theta frequency bands, since these frequency bands are based on exactly the same signals, but analysed differently. The increase in these frequency bands was shown by Baar-Eroglu et al. (2001), but are not exclusive to surprise. This way, only the sensitivity criterion is met. However, an increased activity in the delta and theta frequency bands can support the P300 response. The increase of delta frequency band activity (1-3Hz) was observed 0-500ms after stimulus onset, for the theta frequency band (3-6Hz) the increase was significant 250-500ms after stimulus onset Baar-Eroglu et al. (2001). The frequency bands are obtained from the raw EEG signal, There are no extra requirements on measuring equipment. The frequency bands can be calculated from a single electrode, thus a multiple electrode set-up required from measuring the P300 response will suffice for the frequency band analysis as well.

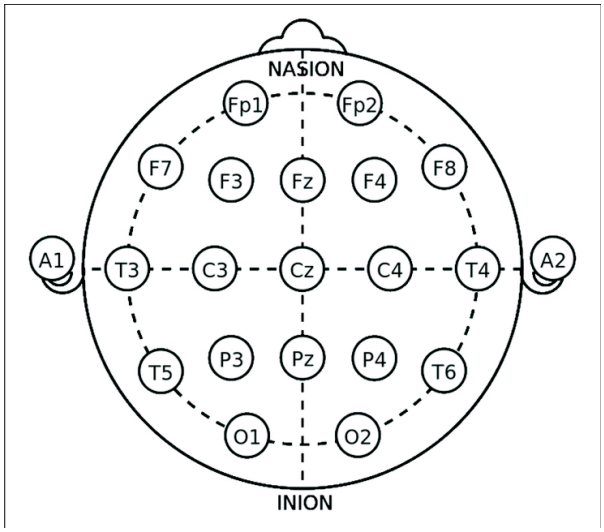


Figure 2.3: Officially supported 10-20 electrode placement pattern for EEG measurements, image taken from Rojas et al. (2018)

Measuring ACC activity directly is of lower value since the P300 component specifically reflects surprise and is easily extracted using the same measurement equipment (EEG) (Mulert et al., 2004).

Pupil dilation related to surprise is slower than pupil dilation after a startling stimulus, reaching a maximum value around 500ms after the stimulus (Kloosterman et al., 2015). thus it is possible to discriminate between the two. Kloosterman et al. (2015) report that pupil dilation represents a global state of the brain during cognitive processing, indicating that pupil dilation is not solely based on surprise. This is supported by Henckens et al. (2009) and Landman (2019) indicating pupil dilation as an effect of stress. Marquart et al. (2015) show the effect of pupil dilation and increased workload. However, it is possible to link pupil dilation to the surprise about behaviourally relevant events during constant lighting (Kloosterman et al., 2015). Pupil dilation is thus sensitive to surprise, but constancy is not always guaranteed (e.g. in stressful situations). Technical requirements on measuring pupil dilation are equal to the ones discussed in the previous section.

Horstmann and Herwig (2014) showed that gaze and dwell time are related to surprising stimuli, with the gaze shifting to the surprising stimulus after 380ms followed by a long (around 600ms) dwelling period on the surprising stimulus. Gaze can be correlated easily to the surprising stimulus if the spatial origin of this stimulus is known and can be measured using eye tracking equipment as discussed in the previous section. Since participants, especially while performing control tasks, are continuously looking around, the time stamp and location of the stimulus need to be known in order to accurately measure if gaze is directed towards the stimulus. In this way, an argument could be provided that gaze doesn't obey to the constancy principle, however since in experimental conditions the stimulus is known, it can still be used as an effective indicator of surprise.

Corrugator muscle activity can be measured using EMG as described in the previous section. The same technical requirements hold. Topolinski and Strack (2015) found that the initial response to a surprising stimulus has a slight increase in corrugator muscle activity up until 1s following the stimulus. After, corrugator muscle activity is dependent on the valence of the stimulus. This indicates that corrugator muscle activity is sensitive to the initial reaction to unexpectedness. It should be noted that frowning can be actively controlled by individuals in other situations, so constancy is not guaranteed. A possibility is to measure increase in muscle activity in relation to the activity before the stimulus, but this is not yet validated. In addition, since Reisenzein et al. (2006) found little increased activity, questions could be raised whether this is as sensitive a parameter for measuring surprise.

Decreased or increased HR is sensitive to surprise according to Blascovich and Mendes (2000), Epstein et al. (1975), and Noordewier et al. (2016). An increase in HR with an onset latency of around 3s was found by Epstein et al. (1975), while Simons et al. (1998) found a latency of around 2s. Decreasing HR starts almost immediately after the stimulus and lasts for around 1s (Simons et al., 1998). HR, although not only influenced by the SNS (Handbook of Psychophysiology, 2016), is characterised by a specific pattern indicating surprise (Simons et al., 1998). A decrease lasting for around 1s followed by an increase in instantaneous heart rate lasting for around 4s can be observed. This makes HR a useful measure for quantifying surprise. Constancy regarding HR is relative to situations. Individuals can control HR to a certain degree, but it is still mostly regulated by the SNS and PNS, especially at shorter time scales such as those for quantifying surprise. HR should thus be supported by other measurements to provide a clearer view. In addition, the time stamp of the stimulus should be included. Measuring HR has been explained in the section on startle.

Since BP either increases or decreases depending on the valence of the stimulus (Blascovich & Mendes, 2000), it does not provide a clear indication of surprise.

2.3.3. Summary

Table 2.1 provides an overview of the above mentioned physiological effects of S&S and their onset latency. In addition, the commonly used measurement techniques for these effects and whether/when they obey to the constancy and sensitivity criteria are provided as well. This table will be used in the next section to discuss the measurements which will be implemented in the sensor suite.

2.4. Choosing measurements

This section deals with the selection of physiological measurements to be implemented in the sensor suite. The measures are selected based on Table 2.1. After selected measures, research is reviewed to obtain insight in whether these measures have already been implemented in flight or simulator settings. Car and car simulator experiments are also included since driving requires approximately the same amount and type of movement. The current literature is reviewed to ensure that implementation of these measurements is possible in flight simulator settings. Three measurements will be selected since this is estimated to fit in the graduation project timeline.

Based on table 2.1, pupil dilation seems to be a measure valid for both startle and surprise. However, for surprise the constancy criterion is not fully met. In addition, blinking and gaze reflect both startle and surprise respectively. This leads to eye tracking looking like a feasible and representative option to measure 4 physiological effects at once. However, an eye tracker is already present in the SRS, so there is no room for development there.

The blinking reflex to a startling stimulus can also be measured using EMG (Blumenthal et al., 2005). Having the EMG equipment installed also allows to measure the trapezius muscle activity. Both measures comply to the constancy and sensitivity criteria as long as the timestamp of the stimulus is provided. Having EMG also allows for measuring the increased corrugator muscle activity related to surprise. While the corrugator muscle activity is is still disputed (Reisenzein et al., 2019), having the measurement equipment in place allows for a validation whether this effect actually occurs, thereby contributing to the current sate-of-the-art. In addition, a way of measuring the orientation towards the startling stimulus can be implemented using EMG. Orientation towards the stimulus is only valid if the participant has not yet habituated to the startling stimulus. However, pilot training scenarios will not be focused on habituation. In addition, the aforementioned measures will be combined with other measures to create a more complete view. As long as the different measurements complement each other, drawbacks of single measurements can be reduced.

Since an increased P300 ERP wave complies to both the constancy and sensitivity criteria, it has to be considered for implementation. Together with the startle eye blink reflex and pupil dilation, this is the only physiological reaction that complies to both of these measures. ICA should be performed on the obtained EEG data to remove artefacts. When having the EEG equipment present, the increase in EEG frequency bands can be obtained with the same set of measurements to support the detection of surprise.

Since HR has a specific pattern related to surprise, ECG can be used to extract surprise effects. Since ECG is then already implemented, the high frequency to low frequency HRV can be measured as well for observing startle. In addition, the increase in HR can then be measured as well for startling stimuli.

The above suggestion on implementations of measurements covers the most constant and sensitive parameters which can be measured without using eye tracking, along with less constant parameters for which supporting measurements are required. This way, most of the effects from table 2.1 are covered with only 3 measurement devices. Following is a discussion of implementation of these devices in the aviation setting or automotive setting.

The aforementioned trapezius muscle activity was measured using EMG in a real world driving task by Healey and Picard (2005) to measure stress levels. Tichon et al. (2014) made use of EMG in a fixed-base flight simulator. Implementing EMG for eye blink measurement has been described extensively by Blumenthal et al. (2005).

Ryu and Myung (2005) used EEG measures for a manual control task. The authors excluded the EEG fragments which were contaminated by eye movement artefacts instead of applying ICA. Rooseleer et al. (2022) uses an off-the-shelf EEG recorder in a moving-base flight simulator. Gibson et al. (2019) used dry-electrode EEG to measure the P300 response is a moving-base flight simulator. EEG in a fixed base simulator was also performed by Johnson et al. (2015).

Bruna et al. (2018) implemented ECG measurements in a fixed base flight simulator, while Hannula et al. (2008) performed the same measurements in a moving base simulator. Healey and Picard (2005) implemented ECG measures in a real world driving task, while Veltman (2002) performed ECG recordings in both simulated and real flight.

Table 2.1: Summarising table of physiological effects belonging to either startle or surprise, their onset latency, measurement techniques and whether they obey the constancy and sensitivity criteria

Startle				
Effect	Latency	Measurement	Constancy	Sensitivity
Eye blink	20-50ms	EMG, eye tracking	if stimulus timestamp provided	yes
Increased trapezius muscle activity	50-400ms	EMG	if stimulus timestamp provided	yes
Orientation towards stimulus	400-450ms	EMG, eye tracking	if not habituated	if not habituated
Increased SCR	1-4s	EDA sensor	if stimulus timestamp provided	yes
Increased BP	4-5s	back-scattered PPG	under controlled conditions	yes
Increased high frequency HRV/low frequency HRV	3-4s	ECG	in controlled conditions	yes, if compared to reference
Increased HR	3-4s	ECG	in controlled conditions	yes, if compared to reference
Pupil dilation	200ms	eye tracking	yes	yes
Surprise				
Effect	Latency	Measurement	Constancy	Sensitivity
P300 ERP increase	300ms	EEG	yes	yes
Increased delta frequency band activity	0-500ms	EEG	no	yes
Increased theta frequency band activity	250-500ms	EEG	no	yes
Pupil dilation	500ms	eye tracking	in controlled conditions	yes
Gaze and fixation on stimulus	380ms	eye tracking	yes	yes
Corrugator muscle activity increase	until 1s	EMG	if compared to reference	yes
Decreased HR	until 1s	ECG	in controlled conditions	yes, if compared to reference
Increased HR	2-3s	ECG	in controlled conditions	yes, if compared to reference

3

Research questions and objectives

Startle & surprise impose risks in the modern cockpit, possibly leading to loss of control. To mitigate the risks of startle & surprise, improved training scenarios can prove to be of great value. In order to produce more effective training scenarios, a measure to quantify S&S needs to be developed. For this reason, a subjective rating scale similar in principle to the NASA TLX scale (Hart & Staveland, 1988) to subjectively quantify the level of S&S is being developed. To improve accuracy of the scale, a set of objective measures can be used to support the subjective ratings during the development process which can serve as a ground truth for quantifying S&S. Since S&S invokes SNS activation, physiological signals can be used to measure a participant's mental state. The research objective is to detect whether pilots are startled or surprised in the SRS by measuring physiological signals. These signals can help in developing a subjective rating scale to quantify startle & surprise. The scale can then be used to further improve pilot training to incorporate startling and surprising events. Physiological signals related to startle and surprise include eye blinks, trapezius muscle contraction, P300 ERP, changes in the EEG delta and theta frequency bands and changes in HR and HRV. These effects can be measured using EMG, ECG and EEG.

3.1. Research questions

The main research question is thus:

> To which degree can surprise and startle be detected by designing and implementing a physiological sensor suite to perform EMG, ECG and EEG measurements?

The subquestions are:

1. Can an increase in trapezius muscle activity, eye blinking, increased HF/LF HRV ratio and increased HR be detected when participants are presented with startling stimuli?
    - Reliable evocation of the physiological signals by startling stimuli
    - Accuracy of the obtained measurements
2. Can the presence of the P300 ERP, increased delta and theta EEG frequency band activity, and the specific HR pattern be observed when participants are presented with surprising stimuli?
    - Reliable evocation of the physiological signals by surprising stimuli
    - Accuracy of the obtained measurements

3. When presented with data that does not provide info about the type of stimuli, to which degree is it possible to extract whether the participant was surprised or startled?

3.2. Hypotheses

In order to answer the research questions stated above, 3 experiments will be performed. The first 2 experiments will serve as a verification of the sensor suite to detect surprise and startle from the obtained measurements. A first experiment serves to verify whether HR, HRV, increased muscle activity and eye blinks can be extracted from the data. The second verification experiment will verify whether the physiological responses to startling and surprising stimuli can be detected. This experiment will be performed under controlled conditions where the participant is not moving. The third experiment will serve as a validation experiment. It will be equal to the second experiment except that no information will be provided in the data t indicate the type of stimulus. This way, the system is validated to detect S&S without knowing stimulus types. In both experiments, the same data will be obtained in order to detect S&S from the participants.

Sub-questions 1 and 2 will be answered by the second experiment. This experiment is solely designed for verifying whether the measured signals can be used to detect S&S. If the outcomes of this experiment are negative, different measurements will need to be considered and the verification experiment will have to be re-performed. Due to literature being unclear at times and due to the complex working of the SNS and PNS, there is no certainty that the measurements will accurately detect responses to startling and surprising stimuli. It is expected that under controlled and stationary conditions:

1. The P300 response will be visible in the EEG traces after a surprising stimulus if enough stimuli are presented during the trial
2. An increase in delta and theta EEG frequency activity will be observed for surprising stimuli
3. Eye blinks will be observable in both the corrugator EMG and frontal EEG traces after a startling stimulus
4. The trapezius muscle will show increased activity after a startling stimulus
5. An initial decrease, followed by an increase in HR will be observed after a surprising stimulus
6. An increase in HR will be observed after a startling stimulus
7. The ratio of high frequency HRV to low frequency HRV will increase after a startling stimulus.

Sub-question 3 will be answered after the third experiment. The purpose of this experiment is to validate whether the sensor suite can reliably detect startling and surprising conditions without knowing the context of the data. It is expected that:

1. Due to not knowing stimulus type, surprise and startle will be detected less frequently compared to the first experiment
2. No effects on EMG
3. No effect on ECG

4

Designing a physiological sensor suite

This section deals with the theoretical design of the physiological sensor suite. First, the general requirements for the sensors are listed. After, different sensor options commonly used in literature will be discussed and compared. The comparison will consist of a comparison against the general requirements and a comparison against the specific technical requirements for each physiological signal to be measured. Based on this comparison, a decision on the sensors to be used will be made.

4.1. General requirements

Each sensor is subject to a list of general requirements to be eligible for incorporation into the sensor suite.

- Galvanic separation between participant and electrical net
- Able to interface with SRS
- Synchronise data to simulation time stamps
- Meet technical requirements for measuring specific physiological effect
- Data sheets, specifications or API available for open development
- Minimally invasive
- Portable
- Economically feasible

First of all, a galvanic separation between the participant and a source of electricity directly connected to the power grid needs to be present. This means that the sensors have to be able to meet their power demands using batteries to avoid direct connection to power sources. Since no direct connection is allowed, the data transfer to SRS and DUECA needs to be performed in a wireless manner as well. Since WiFi is not available on DUECA, wireless communication is performed using the Bluetooth protocol. Currently, only Bluetooth low energy (BLE) is configured, but older versions of Bluetooth can be implemented as well. Another option of getting data into DUECA is streaming data from the sensors wirelessly to an acquisition system, which can then send the data through a physical connection to the simulator. Third, the sensor needs to meet the specific requirements for measuring the physiological effect at hand. These requirements include:

- High sensitivity and resolution
- Wide measurement range to accommodate extreme cases
- Adequate frequency range
- Low noise level
- Hardware to amplify, filter and digitise the raw signals

To be able to develop interfacing software, a clear specification of the communication protocol used by the sensor needs to be provided. If possible, a C++ application programming interface (API) should be available in order to write custom applications for the sensor. Furthermore, the device needs to be minimally invasive for use in a flight simulator and it needs to be portable and comfortable to wear. The device needs to be priced at an economically feasible level. Lastly, an advantage is provided to all-in-one systems: systems that can incorporate all the required sensors, eliminating the need for developing multiple interfaces for different brands. In addition, synchronisation between sensors and DUECA will be less inconvenient if all sensors are made by the same manufacturer and use the same protocols.

4.2. Considered sensors

Four options will be considered for implementation into the simulator. These are established systems incorporating both EMG, ECG and EEG into one system, namely Biopac systems (Biopac Systems, 2023c), Bitalino (r)evolution (Plux Biosignals, 2023) and OpenBCI Cyton (OpenBCI, 2023e). The fourth option for each measurement type consists of a sensor specifically designed for the measurement at hand. These options thus require synchronising the sensors separately and developing an interface to DUECA for each individual sensor. For EMG and ECG, Shimmer sensors will be compared to the other three options (Shimmer Sensing, 2022a, 2022b). Considering wireless EEG sensors, many options are available. Niso et al. (2022) reviewed 44 wireless EEG devices and categorised them according to the amount of channels, data protocol, certification and electrode type, as shown in figure 4.1. This figure indicates the mere size of available options, each with their own advantages and disadvantages. Niso et al. (2022) found that for mobile recording, Emotiv (Emotiv, 2023b), mBrainTrain and ANT Neuro were the most commonly used brands. Since measuring the ERP P300 wave is only performed at 4 electrode sites, 4 input channels for recording EEG suffice.

A similar review by Tajdini et al. (2020) also looked at the application domains of the EEG sensors. For measuring ERP, both the Emotiv and OpenBCI (OpenBCI, 2023e) brands were used most often. This leads to the selection of the Emotiv as the fourth possible option for measuring EEG.

Biopac is a well established medical solution for physiological data acquisition and analysis. A quick search using "Biopac" in Google Scholar yielded 48400 results. It has been proven in experimental physiological studies (Leuchs et al., 2019). The company provides a range of wearable wireless sensors, called BioNomadix, suitable for measuring EMG, ECG and EEG (Biopac Systems, 2023a). Acquisition of these signals is performed by an MP160 acquisition system (Biopac Systems, 2023d) which is then connected to a host computer. A price estimate was provided by the manufacturer to be around €10000. Proprietary software is provided with the system that can be used to analyse and log the data. Data sheets are publicly made available (Biopac Systems, 2023a). In addition, the company offers C++ API's to directly control the acquisition and transportation of data via the TCP/IP protocol, thus facilitating custom applications and data synchronisation. The MP160 system has 16 input channels, ensuring expansion to more physiological sensors is possible. In addition, the input

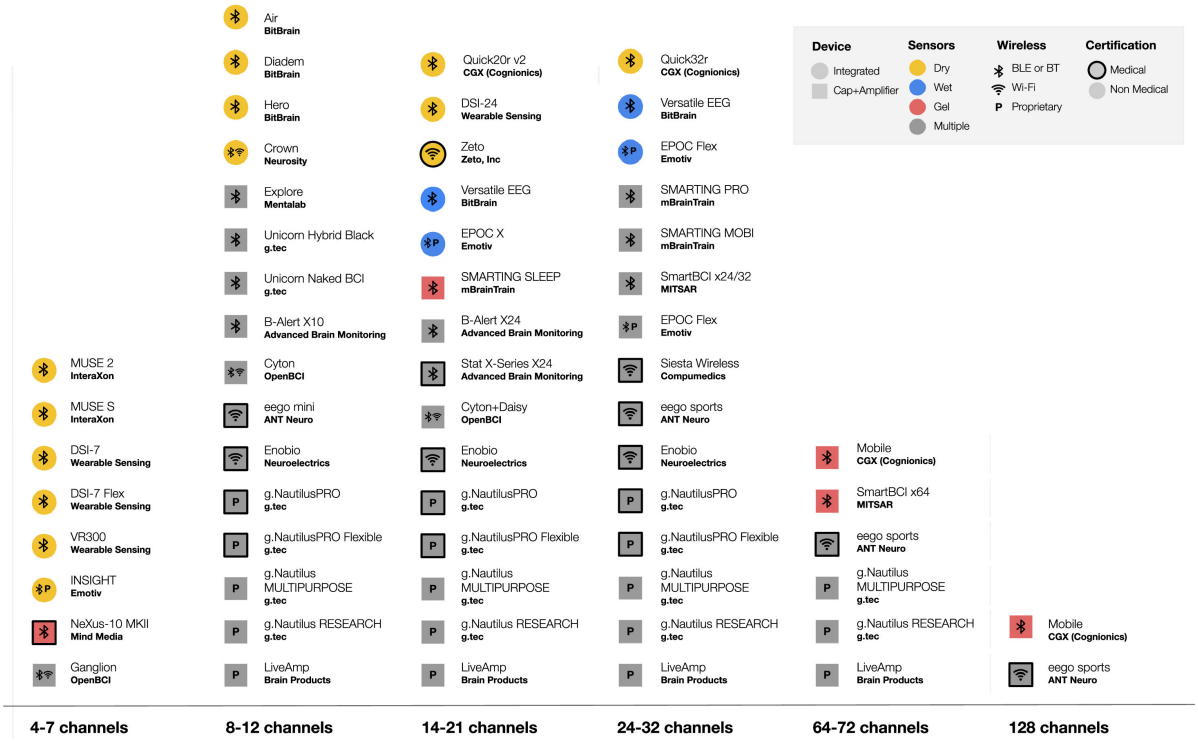


Figure 4.1: Selection of wireless EEG sensors categorised by Niso et al. (2022) according to the amount of channels, communication protocol, certification and electrode type. Figure taken from Niso et al. (2022)

channels can be used to attach third-party sensors as well. A schematic representation of the BioNomadix sensor combined with the MP160 acquisition system is shown in figure 4.2.

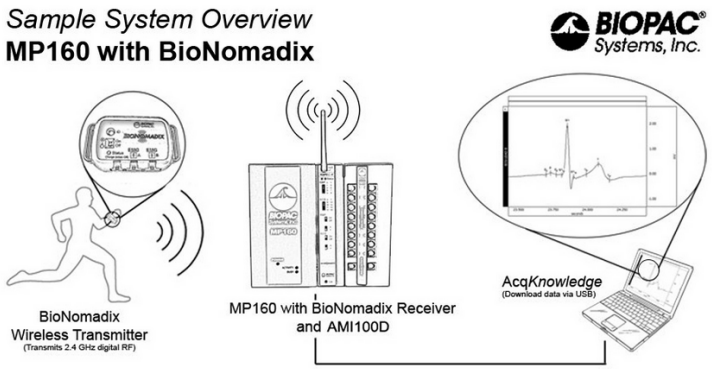


Figure 4.2: Schematic representation of the Biopac BioNomadix wireless sensor combined with the MP160 acquisition system (Biopac Systems, 2023a)

Bitalino is an upcoming, low cost and open source development platform for physiological measurements. A full system costs around €300. It is designed for rapid development of custom education and laboratory applications. Signal quality has been bench-marked to the Biopac system and Biosignalplux system in Batista et al. (2019), Batista et al. (2017). Results showed a good comparison between the gold standards and the Bitalino (r)evolution board when comparing relevant physiological metrics. The board communicates using the BLE protocol and has C++ API's available for implementing custom applications (pluxbiosignals, 2019). The board has 4 main 10 bit input ports for physiological sensors. Additional lower resolution ports are available as well. Various sensors are available, including EEG, ECG and EMG.

Third party analog sensors can be connected to the board as well. Much like Biopac, Bitalino provides proprietary software to stream, log and analyse the data. An example layout of the (r)evolution board without sensor leads or battery is depicted in figure 4.3.

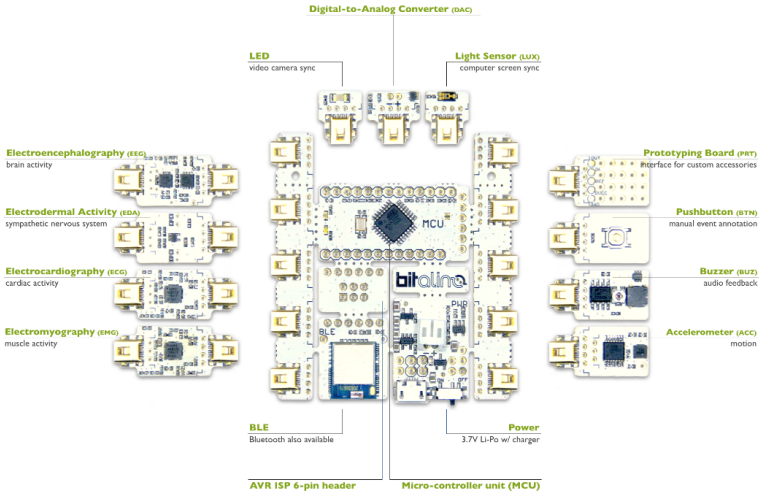


Figure 4.3: Bitalino (r)evolution overview, excluding sensor leads and battery (Plux Biosignals, 2016a)

OpenBCI has a similar design mentality to Bitalino, with a larger focus on EEG measurements. However, it also provides ECG and EMG functionality. The Ganglion is an open-source research board with 4 input channels, all sampled at 200Hz. Prices for a complete EEG setup lie at around €800. The board is equipped with a BLE module and extensive documentation is provided for custom development (OpenBCI, 2022b). It has been effectively used for measuring P300 responses (Frey, 2016; Joshi et al., 2022). The system is not meant for medical use, but focused on quick prototyping and experimenting. A lot of room for customisation is incorporated into the design.

In theory, 4 input channels suffices, since 2 EMG measurements, 1 ECG measurement and 1 EEG measurement (on either the PZ, Nz or Cz scalp locations) needs to be performed. However, for ensuring EEG redundancy and possible future extension of measurements, more channels may be required. The second option from the OpenBCI brand is the Cyton board (OpenBCI, 2023d). It contains 8 input channels and also supports active EEG electrodes. A setup with standard electrodes costs around €1100. the same focus of rapid prototyping and custom development as for the Ganglion board is adopted. In the following sections, the Cyton board will be compared against the other options, due to having more channels to future-proof the design.

The shimmer platform has been used in research as well (e.g. Mohd Azli et al. (2019) and Rooseleer et al. (2022)) and provides a middle ground between Biopac and Bitalino in terms of price range. A single sensor costs around €500. As with Biopac and Bitalino, Shimmer provides its own software. No explicit C++ API's are provided, but a full specification about the data sent over via Bluetooth is available (Shimmer Sensing, 2017b). The system uses the RFCOMM Bluetooth communication framework, not BLE. Still, it is possible to integrate this system in DUECA. Each Shimmer sensor is specifically designed for measuring 1 physiological signal. The sensors are portable, wireless and battery powered. An example implementation of the EMG unit can be seen in figure 4.4. The ECG sensor has the exact same footprint thus no separate image is provided.

Emotiv is a consumer grade portable EEG device. It is less flexible in design and has fixed electrode placements. The Epoc X model has been used in P300 research and contains 14 channels, with a base price of around €850 (Emotiv, 2023a). Electrodes are not placed on the

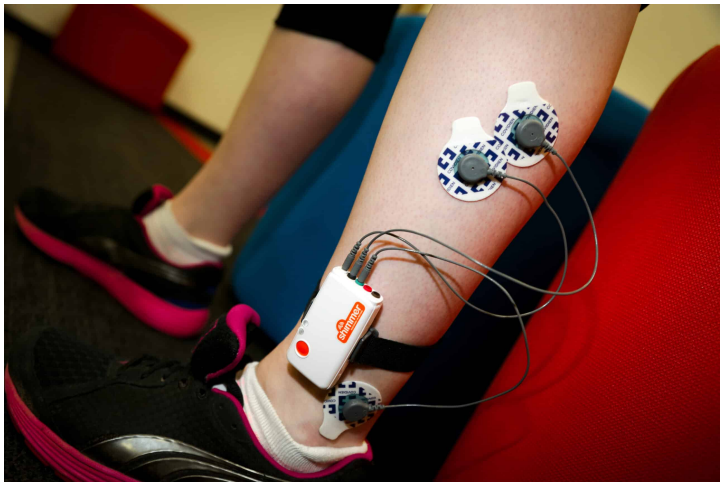


Figure 4.4: Shimmer EMG sensor example implementation. Picture taken from Shimmer Sensing (2022b)

Fz, Cz and Pz scalp positions, but positions near the centre line and on the front of the scalp. API's are available for acquiring the raw EEG data for $\$29/month$. Data is transmitted using the BLE protocol. Proprietary software is available for configuring the sensors and acquiring data. Figure 4.5 shows the Epoc X headset.

Figure 4.5: Emotiv Epoc X EEG headset (Emotiv, 2023b)

4.3. EMG comparison

Technical requirements for measuring EMG are documented by Blumenthal et al. (2005) and listed below.

1. A wide frequency range to capture both low-frequency muscle activity and high-frequency noise, ideally from 0-500 Hz.

2. Low noise level to minimise interference with the muscle signals, ideally less than $1\mu V$ RMS in the 10-500Hz frequency range.

3. High input impedance to reduce loading effects on the muscle, greater than $100M\Omega$.
4. High common mode rejection ratio (CMRR),ideally greater than 80dB.
5. High signal-to-noise ratio (SNR), ideally greater than 50dB.
6. Sampling rate of 1000Hz minimum
7. ADC resolution such that μV changes can be measured

Specifications for all three sensor options are compared against these requirements in table 4.1. The data for each sensor is obtained from the sensor data sheets provided by the manufacturers (Biopac Systems, 2023a, 2023d; Plácido da Silva et al., 2014; Plux Biosignals, 2016a, 2016b; Shimmer Sensing, 2017a, 2017c).

Table 4.1: Comparison of the Biopac, Shimmer, OpenBCI and Bitalino sensors against EMG requirements proposed by Blumenthal et al. (2005). Data obtained from manufacturers (Biopac Systems, 2023a, 2023d; OpenBCI, 2022b; Plácido da Silva et al., 2014; Plux Biosignals, 2016a, 2016b; Shimmer Sensing, 2017a, 2017c; Texas Instruments, 2017)

	Requirement	Biopac	Shimmer	OpenBCI	Bitalino
Frequency range	0-500Hz	5-500Hz	0-8.4kHz	>24kHz	25-480Hz
Noise level	$< 1\mu V$ RMS for 10-500Hz	$1.5\mu V$ RMS for 5-500Hz	$5mV$	$1\mu V$	NS
Input impedance	$> 100M\Omega$	$100M\Omega$	NS	$1G\Omega$	$7.5G\Omega$
CMRR	$> 80dB$	$90dB$	NS	$120dB$	$86dB$
SNR	$> 50dB$	$65dB$	NS	$121dB$	$55.72dB$
Sampling rate	$> 1000Hz$	$2000Hz$	until $8000Hz$	$250Hz$	$1000Hz$
Resolution	μV range	$1.5\mu V$	$0.28\mu V$	$0.5\mu V$ for gain=1	$3.2\mu V$

No SNR is provided for the Shimmer system; However Burns et al. (2010) compared the SNR in a different metric to a well established Grass amplifier and observed the SNR to be higher. González-Mendoza et al. (2018) reported a SNR of 22.7dB for the Shimmer EMG sensor. In addition, no CMRR or input impedance or was found for the Shimmer system. No noise level units were specified in the documentation for the Shimmer system, so Volts was assumed as a unit. This lack of data makes for a difficult comparison between the avialable sensors.

The same goes for the Bitalino board, for which no noise level is provided.

In general, the Bitalino board, even at a fraction of the price, seems to perform almost equally well as the Biopac system. In terms of frequency range, the Bitalino board does not live up to the requirement, but since the system will only be used to measure an increased value in muscle activity with respect to a baseline, it should not provide many problems. Due to a lot of lacking data, the Shimmer device is difficult to compare. However since it has been used successfully in research, we assume the data quality suffices. The Biopac system is medically certified and thus guarantees an accurate data output. If very accurate measurements should be performed, this would be the device of choice.

In terms of adaptability and interfacing, the Bitalino board is a clear winner. It comes with C++ API's and leaves a lot open for custom development. Transmission is performed via

BLE, allowing for easy integration and synchronisation into DUECA. The Biopac system can also be integrated into DUECA, but using a wired connection between the MP160 acquisition system and the simulator. This makes the system less suitable for use in the different TU Delft simulators and brings with it the need to develop a completely new interface based on TCP/IP to control the complete acquisition system. In addition, the large footprint of the MP160 system makes for a sub-optimal integration in the simulator due to space constraints. For acquiring the raw Shimmer output, a fully custom interface will have to be developed, since only the data packet structure is specified in the user manual. The unit works with RFCOMM, an older Bluetooth protocol. Many unknowns regarding communication with this sensor will have to be answered during development of this interface. No C++ API is made available.

4.4. EEG comparison

Technical requirements for measuring the P300 potential using EEG are not as well defined as EMG. In this section, the recommendations presented by Duncan et al. (2009), Habibzadeh Tonekabony Shad et al. (2020), Halford et al. (2016), and Picton et al. (2000) are used. The guidelines differ in some aspects. In these cases, the most stringent guideline is chosen.

1. Sampling rate of 256HZ
2. ADC resolution of 16 bits
3. CMRR of 100dB
4. Resolution of $0.05\mu V$
5. A low noise level $< 1\mu V$ in the 0.5-100Hz range
6. Inter-channel cross-talk should be $< 40dB$
7. Input impedance $> 100M\Omega$ at 50/60Hz

Table 4.2 provides an overview of how the Biopac, Bitalino and TODO systems stack up against these requirements. Again, all data is obtained using information published by the manufacturer (Biopac Systems, 2023b, 2023d; Emotiv, 2023a, 2023b; OpenBCI, 2022b; Plácido da Silva et al., 2014; Plux Biosignals, 2016a, 2018; Texas Instruments, 2017).

In terms of performance, the Biopac system outperforms all other options in almost all categories. The Bitalino board has a low ADC resolution, but due to its low input range ($\pm 39.4\mu V$ Plux Biosignals (2018)), a good voltage resolution can still be obtained. However, the EEG signal can be into the mV range due to e.g. motion artefacts (Niso et al., 2022), so the Bitalino board can be subject to saturation in conditions where the participant is moving. In comparison, the Emotiv system has an input range of $8.4mV$ (Emotiv, 2023a), the OpenBCI $375mV$ (for a gain of 24, see Texas Instruments (2017)) and the Biopac system $2mV$ (Biopac Systems, 2023a). These ranges are more suitable for measuring EEG without saturation. This leads to believe that the Bitalino board might not prove to be accurate enough in scenarios where the participant is moving.

The same arguments in terms of customisation for the Bitalino board and Biopac system hold. Comparable to the Bitalino board, the OpenBCI is suitable for custom development. It uses BLE and has extensive documentation for custom development and raw data acquisition. The interface for OpenBCI and Bitalino is likely to be very similar, given that they both use the same protocol. Emotiv, as mentioned earlier, does provide API's for raw data acquisition using the BLE module. Whether this API is based on C++ is still unknown at the time of

Table 4.2: Comparison of the Biopac, OpenBCI, Emotiv and Bitalino sensors against EEG requirements proposed by Duncan et al. (2009), Halford et al. (2016), and Picton et al. (2000). Data obtained from manufacturers (Biopac Systems, 2023a, 2023d; Emotiv, 2023a, 2023b; OpenBCI, 2022b; Plácido da Silva et al., 2014; Plux Biosignals, 2016a, 2018; Texas Instruments, 2017)

	Requirement	Biopac	OpenBCI	Emotiv	Bitalino
Sampling rate	$> 256Hz$	$2000Hz$	$250Hz$	$256Hz$	$1000Hz$
ADC resolution	16 bits	16 bits	24 bits	16	10 bits
CMRR	$> 80dB$	$90dB$	$120dB$	NS	$100dB$
Resolution	$0.05\mu V$	$0.03\mu V$	$0.02\mu V$ if gain=24	$0.12\mu V$ in 16 bit mode	$0.077\mu V$
Noise level	$< 1\mu V$ RMS for 0.5-500Hz	$0.2\mu V$ RMS for 0.1-100Hz	$0.16\mu V$ at $60Hz$ if gain=24	NS	NS
Interchannel cross-talk	$< 40dB$	NS	$-121dB$	NS	NS
Input impedance	$> 100M\Omega$ at 50/60Hz	$2M\Omega$	$1000M\Omega$	NS	$> 100G\Omega$

writing. If however, this turns out to be the case, then the same interface can be used for the OpenBCI, Emotiv and Bitalino sensors.

In terms of unobtrusiveness, the Biopac, Bitalino and OpenBCI platforms all allow for free electrode placement on the desired scalp positions, while the Emotiv system has electrodes fixed in place. The design of the Emotiv system removes the ability to wear headphones used for realistic flight simulations. Having 14 channels is however beneficial if future work aims to measure more than the single P300 potential.

The Bitalino EEG sensor has only one channel, thus if measurement from multiple regions should be performed, multiple EEG sensor units will have to be connected, reducing the number of available ports left. In the case of measuring the P300 response, a single channel placed on either the Pz, Nz or Cz scalp location suffices (Duncan et al., 2009). Every BioNomadix EEG sesor provided by Biopac contains 2 EEG channels. However, since the MP160 has 16 input ports, multiple sensors can be used without compromising other measurements. The OpenBCI Ganglion board has 4 input channels, which is sufficient to measure the P300 response at both the Pz, Nz and Cz locations.

4.5. ECG comparison

Technical requirements for implementing ECG are either outdated or vaguely specified. The recommendations listed below are taken from Bailey et al. (1990), Guerrero and Spinelli (2022), Malik and Camm (1990), Rinold and Porges (1997), Willems (1985), and Wolf et al. (1976). Conflicting recommendations were solved by taking the recommendation published more recently. Extensive technical requirements are provided using IEEE/IEC standards (Diagnostic Electrocardiographic Devices, 1996; Medical electrical equipment - Part 2-25: Particular requirements for the basic safety and essential performance of electrocardiographs, 2011; Medical electrical equipment Part 1-11: General requirements for basic safety and essential performance Collateral standard: Requirements for medical electrical equipment and medical electrical systems used in the home healthcare environment, 2015), however this information is not publicly accessible without payment.

1. Frequency range of 0.67Hz to 40Hz
2. Sampling rate of 500Hz or higher
3. Resolution of $10\mu V$
4. Noise level of $< 10\mu V$ RMS
5. ADC resolution of 10 bits
6. CMRR of 90dB
7. Input impedance $> 10M\Omega$

Table 4.3 provides an overview of the sensor specifications obtained from the manufacturer data compared to the technical requirements specified above (Plácido da Silva et al., 2014; Plux Biosignals, 2016a, 2020).

Table 4.3: Comparison of the Biopac, Shimmer and Bitalino sensors against ECG requirements proposed by Bailey et al. (1990), Guerrero and Spinelli (2022), Malik and Camm (1990), Rinold and Porges (1997), Willems (1985), and Wolf et al. (1976). Data obtained from manufacturers (Biopac Systems, 2023a, 2023d; OpenBCI, 2022b; Plácido da Silva et al., 2014; Plux Biosignals, 2016a, 2020; Shimmer Sensing, 2017c, 2018; Texas Instruments, 2017)

	Requirement	Biopac	Shimmer	OpenBCI	Bitalino
Frequency range	0.67-40Hz	0.05-150Hz	0-8.4kHz	>24kHz	0.5-40Hz
Sampling rate	$> 500Hz$	$2000Hz$	until $8000Hz$	$250Hz$	$1000Hz$
Resolution	$< 10\mu V$	$0.2\mu V$	$0.28\mu V$	$0.5\mu V$ worst case	$2.9\mu V$
Noise level	$< 10\mu V$ RMS	$0.9\mu V$ RMS	NS	$1\mu V$ RMS worst case	NS
ADC resolution	10 bits	16 bits	23 bits	24 bits	10 bits
CMRR	$> 90dB$	$90dB$	NS	$120dB$	$86dB$
Input impedance	$> 10M\Omega$	$1G\Omega$	NS	$1G\Omega$	$7.5G\Omega$

All sensors perform very similar. The noise level for the Shimmer and Bitalino board are not provided. As with the EMG specifications, not a lot of information regarding the Shimmer specifications can be obtained. This makes for a difficult comparison. The Bitalino board seems to perform well. However, the input voltage range is specified as $\pm 1.5mV$ (Plux Biosignals, 2020). According to practical recommendations provided by Nmcová et al. (2016), this range can prove to be inadequate for some participants, resulting in trimming of the R-peak. The problem can be solved by placing the electrodes further from the heart (at the wrists), but this increases sensitivity to motion. The noise level was evaluated by Nmcová et al. (2016) as well, and proved to be satisfactory to obtain all required information form the signal, even during movement.

4.6. Discussion

From the previous sections, it is clear that every option discussed has its advantages and disadvantages. This section will formulate a final decision, based on the technical specifications mentioned above, and on the requirements for this specific project.

Purely specification based, the Biopac system proves to be a clear winner. However, the high price point, large footprint and the necessity to use a wired connection with SRS prevent it from being the versatile sensor suite this project desires. The system has a clear focus on highly accurate measurements in a controlled environment to be used for medical diagnoses. Since this project only needs a measurement accuracy sufficient to detect changes in levels or the presence of physiological signals, using medical grade equipment would be excessive.

The Shimmer ECG and EMG sensors are well used in research; However, there is little data available about the technical specifications. Since it is such a well-established system, it s assumed the technical specifications suffice for the purpose of this sensor suite. The bigger problem with the Shimmer sensors is the RFCOMM wireless transmission and the lack of adequate tools for supporting development of custom applications. This makes integrating the sensors in DUECA a difficult and unpredictable task.

The Emotiv system has sufficient technical capabilities for measuring the P300 response. In addition, there are possibilities to extract raw data using the BLE protocol. The Epoc X headset has a fixed electrode placement, which reduces the options for measuring only a few channels and on locations different from the fixed electrode positions. In addition, the headset has a large footprint which reduces the possibilities for the participants to wear headphones, commonly used in the SRS.

The Bitalino system has a economical price tag, and adequate technical specifications. In addition, there exist support for custom applications and data is transmitted via BLE. One disadvantage from this system is the saturation limits on the EEG sensor. When the participant is moving, this might result in faulty EEG readings. The same goes for the ECG sensor, which can attenuate R-peaks of high input voltage.

The OpenBCI Cyton board is also economically priced with lots of capabilities in terms of custom development and uses the BLE protocol. Technical specifications suffice for almost every measurement. In addition, it provides more input channels than the Bitalino board. However, the sampling rate of 250Hz can be too low for obtaining valid EMG measurements, since most of the EMG information lies in to 20-350Hz range (Juárez et al., 2022). The limited sampling rate is due to the data being transmitted using BLE. Since the main goal of the EMG sensor for this project is measuring increased muscle activity, a high sampling rate might not be necessary. The 250Hz sampling rate has been used often (Camilo et al., 2020; Cao et al., 2019; Cassani et al., 2020; Chen et al., 2021; Ma'as et al., 2017; Ugarte et al., 2019; Valderrama et al., 2021).

To overcome the limited sampling rate, OpenBCI has developed a WiFi board which allows for sampling rates up until 16kHz (OpenBCI, 2022c). This product is however not for sale from the official manufacturer due to some setups encountering issues and packet loss. It is possible to construct the WiFi shield from the diagrams (Olofsson, 2019) or to buy it from an external supplier. It should be noted that using the WiFi shield does not provide a guarantee for success. Juárez et al. (2022) has used the WiFi shield successfully. Anther option of mitigating the low sampling rate problem is by connecting a MyoWare sensor to the input ports of the OpenBCI board, instead of directly connecting electrodes. This approach was taken by Peterson et al. (2020). The MyoWare sensor has the ability to directly output the EMG signal envelope, circumventing the need for a high sampling rate (MyoWare, 2022).

Even though the sampling rate might prove to be a limiting factor for measuring EMG, the OpenBCI board still comes out as favourite of the options considered. The EEG capabilities are similar to medical grade hardware (Cassani et al., 2020; Frey, 2016; Lakhan et al., 2019; Rashid et al., 2018), as are the ECG capabilities (Burma et al., 2021). If the EMG sampling frequency does prove to be inadequate, a MyoWare sensor can be connected of a WiFi shield can be bought/constructed. Another option is to use a Bitalino board for EMG measurements,

since it does not have a tendency to saturate for EMG measurements. In addition, this would expand the capabilities of the sensor suite to incorporate e.g. EDA sensing as well. Interfacing the Bitalino is also performed using BLE, so minimal changes to interfaces are expected.

Active electrodes ensure a higher SNR and help to suppress cable artefacts and reduce the need to shield cables (Habibzadeh Tonekabony Shad et al., 2020). If high noise values are observed, active electrodes could be an option to mitigate these effects. Dry electrodes are more susceptible to artefacts but increase user comfort (Habibzadeh Tonekabony Shad et al., 2020).

4.7. Software integration

The Cyton board uses a BLE dongle which plugs into the computer and acts like a serial port. This is done in order to achieve higher data rates than using standard BLE (OpenBCI, 2022a). Communication with the Cyton board can be established using the BrainFlow C++ API (BrainFlow, 2019a). No low level communication protocols need to be coded if using this API, only arguments specifying the board parameters need to be provided (BrainFlow, 2019b). The API has to be included in the DUECA module files and from there, direct calls can be made to use the functions provided by BrainFlow. BrainFlow also provides functions for signal filtering, processing and machine learning techniques (BrainFlow, 2023). Software integration will thus consist of building the API and including it in the relevant DUECA files. From there, data can directly be streamed from the sensor board.

4.7.1. Dependencies

To interact with the virtual serial port, an FTDI driver compatible with the operating system needs to be installed. On Ubuntu, this driver should be pre-installed out of the box. In VirtualBox, the extension packs corresponding to the used version of VirtualBox need to be installed as well. Checking whether the driver has been properly installed can be performed by plugging in the dongle and running:

```
$ lsusb
```

. If the FTDI device shows up, the driver should be properly installed. In addition, the current user needs to have permission to interact with the serial port and should be added to the dialout group (OpenBCI, 2023c). To increase data smoothness, the latency_timer file, should be adjusted to have a value of 1. When the dongle is plugged in, the file is located at:

```
$ /sys/bus/usb-serial/devices/ttyUSB0/latency_timer
```

. To verify correct driver instalment and permissions, the OpenBCI GUI can be downloaded and run with the Cyton board to verify that data can be transmitted between the board and the dongle, and the dongle and PC. Instructions on downloading and installing the GUI can be found at OpenBCI (2023b). The steps inf OpenBCI (2023a) should be followed to ensure the board is communicating correctly.

4.7.2. Building and using the BrainFlow API in DUECA

The BrainFlow source is located at BrainFLow (2023). The API can be built in the cloned repository using:

```
mkdir build && cd build
cmake ..
make -j 8
```

Implementing the API in DUECA is achieved by adding the correct statements in the CMakeLists.txt file of the module using the API. An example can be found at Bogaerts (2023). Building the application using cmake should be performed with an extra parameter set:

```
// in build folder of project
cmake -DCMAKE_PREFIX_PATH=<path_to_brainflow>/installed ..
```

The API can then be used in the same way as the examples provided by BrainFlow (BrainFlow, 2019a).

5

Verification & validation

This section describes the steps to be taken in order to verify and validate the sensor suite. First, the verification, comprising 2 steps, will be discussed. This is followed by a discussion on validation of the constructed sensor suite.

5.1. Verification

To verify the correct working of the Cyton board, and whether it is able to detect the physiological effects described in literature, the verification will consist of 2 stages. The first stage consists of checking whether the heart rate, heart rate variation, differences in muscle activity, and EEG frequency bands can be obtained from the data output of the board. After, the second stage will consist of performing a S&S experiment. Data obtained from this experiment will be used in order to verify if the physiological effects described in literature can be measured using the Cyton board.

A first experiment to verify the correct working of the sensors will be performed. This experiment will take place under controlled circumstances, with the participant sitting still on a chair and looking at a computer screen. Since the sensor suite will be used in conditions where the participant is looking around, the experiment will be performed with eyes opened. This also allows to recognise eye blink artefacts in the EEG data. During the experiment, startling and surprising stimuli will be presented.

5.1.1. Experiment 1

Since the first experiment aims at verifying whether HR, HRV, muscle tension and EEG frequency bands can be effectively computed, the experiment does not have to be complex. Data only has to be acquired and analysed for a single participant and stimuli do not need to be present.

Set-up

The participant will be in a seated position with electrodes attached as described in the next section. Once in a while, the participant will flex the trapezius muscle, blink, and take some deep breaths. Exact times of these actions are not provided since the goal is purely to check the signal quality, not to evaluate the effects of S&S.

Data acquisition

EMG data will be obtained from the lower obricularis oculi muscle and trapezius muscle. The electrodes will be placed according to the conventions in Stegeman and Hermens (2007). No MyoWare sensor or WiFi shield will be connected to the Cyton board.

For measuring ECG, a single Einthoven lead II configuration will be used, with electrodes placed on the right clavicula and left musculus pectoralis major (Nmcová et al., 2016). The reference electrode will be placed under the left clavicula. However, if more suitable electrode locations are found, these can be used as well, such as the proposed placement by Burma et al. (2021)

EEG measures will be obtained from the Fz, Cz and Pz scalp locations to measure the energy contained in the delta and theta frequency bands. Reference will be placed on the earlobe, either on the A1 or A2 location from figure 2.3 (Duncan et al., 2009). To measure eye blinks, an additional electrode will be placed on the forehead of the participant, on either the Fp1 or Fp2 locations from figure 2.3.

All data will be acquired at a sample rate of 250Hz. Data will be sent using BLE to the OpenBCI Cyton dongle connected to the host computer running the simulation. Starting data acquisition and controlling the board will all be performed in DUECA. Inside DUECA, the obtained data will be logged and written to an .hdf5 file. This file format is compatible with both Python and MATLAB.

Data analysis

Since no startling or surprising stimuli will be presented, no P300 response will be elicited. Thus, data analysis concerning this P300 response will be omitted here. The analysis will be performed in Python since it supports BrainFlow and other often used tools for physiological data analysis.

EMG data can be analysed in both the time domain and frequency domain. The first method consists of applying band-pass filters and a 50Hz notch filter. The lower filter threshold is recommended to be placed at 28Hz by Blumenthal et al. (2005). The higher frequency threshold was put at 500Hz. Next, the signal is rectified. After, either smoothing of integration can be performed. Digital integration is achieved by multiplying the average signal value over a certain period with the period duration. Smoothing can be achieved by either using an infinite impulse response filter configured as a digital RC-circuit, or by using a finite impulse response boxcar or variable weight filter. Integrating the signal is more suitable for comparing average muscle activity.

In the frequency domain, analysis exists of performing a Fourier transform of a certain time interval and integrating the obtained power spectral density function to get an indicator of overall muscle activity. This comparison enables to recognise muscle activation after stimulus onset and will be used for this experiment.

If the OpenBCI Cyton board turns out to have an insufficient sampling rate, the MyoWare sensor can be connected to an input pin. This sensor then outputs the smoothed EMG signal, which can be sampled at a lower frequency (Blumenthal et al., 2005).

Both HR and HRV rely on the detection of R-peaks in the ECG signal. With the R-peaks detected, computing instantaneous HR simply consists of dividing 60 seconds by the time difference between two consecutive R peaks. Considering HRV, the RMS and standard deviation of RR intervals' successive differences are commonly used time domain features (Alberdi et al., 2016). In the frequency domain, the ratio of the high and low frequency HRV is commonly used. By performing a Fourier transform on the HRV data, this ratio can be calculated.

Calculating the energy in the EEG delta (1-3Hz) and theta (3-6Hz) frequency bands after stimulus onset is achieved by Fourier transforming the signal 500ms after a fictive stimulus, and comparing it against a baseline before the stimulus.

Participants

Since this experiment's goal is to validate signal quality, the author will be the participant.

5.1.2. Experiment 2

This experiment serves to verify whether the physiological effects described in section 2.4 can be extracted from the data obtained by the Cyton board in a S&S setting. For this experiment, startling and surprising stimuli will be provided to the participants.

Setting

Participants will be seated in front of a digital screen, connected to speakers. EMG, ECG and EEG measurements will be taken to detect changes in trapezius muscle activity, eye blinks, changes in HR and HRV, the P300 surprise ERP and changes in EEG frequency band activity. The participant will look at the screen during the whole experiment. Startling and surprising stimuli will be presented. Data will be acquired using the OpenBCI Cyton board and directly sent to a DUECA simulation. The obtained data will be analysed using BrainFlow and Python to detect startle and surprise. In this experiment, the stimulus timestamps will be known when analysing the data.

Stimuli

An oddball task will be performed to elicit the surprise reaction. This task is commonly used in literature to generate P300 responses (Baar-Eroglu et al., 2001; Duncan et al., 2009; Reisenzein et al., 2019). However, for the verification experiment, a more general surprise inducing stimulus needs to be used. Gross and Levenson (1995) used film fragments to elicit surprise. However after watching the specified fragment, the definition of 'surprise' corresponds to the definition of 'startle' used in this project. Since surprise is only evoked for task-relevant stimuli, eliciting surprise without the participants performing a control task can be difficult. Meyer et al. (1991) used a visual task, whereby the surprising stimulus comprised a word written in a different colour. Instead of an oddball task with 2 different stimuli, Simons et al. (1998) used 3 stimuli. 2 of the stimuli were rare, and of these two, one was specified as the target. The response to the other rare stimulus can be seen as a surprise response since the stimulus is task-relevant and unexpected. The stimuli used for eliciting surprise will be based on this principle.

Participants will be sitting in front of a screen and are told to categorise the duration of target stimuli as "long" or "short", with the first target stimulus specified to be "long". Stimuli will consist of coloured dots, with green dots being the target stimuli and red and blue dots being non-target stimuli. Blue dots will be used as rare non-target stimuli. The participant is not informed about the existence of the blue dots. So when a blue dot appears, a surprise reaction will be evoked.

Startling stimuli will be purely auditory an consist of broadband noise between 20Hz and 20kHz at an intensity of 105dB sound pressure level, following the guidelines of Blumenthal et al. (2005). Rise time will be kept to a minimum and the duration will be around 50ms. Controlling background noise will not be possible, but measurements of background noise will be made to increase reproducibility of the experiment. The sounds will be outputted using speakers to minimise headphone interference during this verification experiment.

The startling stimuli will be presented during the surprise experiment. This makes the stimulus unexpected and not task relevant, thereby inducing startle.

Participants

Since the startle & surprise reactions are independent of age, demographic or other, participants only have to fulfil the requirement of not having any cardiovascular diseases or epilepsy.

Data acquisition

Data acquisition is equal to the acquisition system of experiment 1. For this experiment, the P300 response will be measured in addition to the EEG frequency bands, but no changes to

the acquisition system should be made for this. When stimuli are presented, the timestamp and type of stimulus will be provided along with the data such that analysis can be performed on the data segments containing the startle and surprise reactions.

Data analysis

Again, data anlysis will be similar to the first experiment, with the exception that the measures will now be compared to a baseline before stimulus onset. In addition, the P300 ERP will be measured. The ERP response can be obtained from a single trial, as long as enough stimuli are provided. Since the P300 response is small in comparison to background EEG signals, the multiple stimuli events are used to average out the responses, thereby eliminating the EEG background noise (Blankertz et al., 2011). Obtaining the ERP response is done in four steps:

1. Pre-processing: The first step in analysing EEG data is to pre-process the raw data to remove noise and artefacts that can obscure the underlying EEG signal. This can include filtering the data to remove high-frequency noise (e.g., using a high-pass filter), removing large-amplitude artefacts (e.g., blinks, muscle activity), and down-sampling the data to a suitable sampling rate for analysis.

2. Epoching: The next step is to cut the continuous EEG data into short segments, called epochs, around the time of the event of interest (e.g., stimulus presentation). This involves selecting a time window around the event and extracting the EEG data within that window for each trial.

3. Baseline correction: After the data has been epoched, the average activity in a pre-event baseline period is subtracted from each epoch to remove ongoing activity and baseline shift. This helps to ensure that changes in the EEG signal are related specifically to the event of interest and not to changes in the ongoing EEG activity

4. Averaging: In this step, the individual epochs are combined to obtain a single, representative ERP response. This is done by taking the mean of the EEG data across all epochs and electrodes for each time point. Averaging helps to increase the signal-to-noise ratio by reducing the impact of random noise on the ERP response.

For obtaining a more robust estimate, multiple trials can be performed. If not enough stimuli are present, signal denoising using wavelet transforms can be used. The number of stimuli needed to reliably detect the P300 response is thus a parameter that will be estimated by this experiment.

5.2. Validation

Validation will be performed by using the same experiment. However, no information on what the stimulus is will be provided along with the data. This experiment's goal is to validate whether S&S can be detected without knowing the type of stimulus. Information on whether a stimulus has occurred will be provided in order to be able to segment data, but no information on the type of stimulus (e.g. rare non-target) will be provided. The setting, stimuli, participants and data acquisition/analysis will be equal to the second verification experiment.

6

Conclusions

Due to the possible negative effects, startle & surprise has been given a lot of attention lately.

The cognitive effects resulting from S&S, are among others, acute increase of stress, freezing, losing situational awareness, impairment of information processing, and disorientation. These cognitive impairments can have disastrous effects such as loss of control. To prevent these dangerous situations, improved pilot training scenarios are being developed along with a subjective rating scale to represent the level of S&S of pilots.

Startle & surprise is also characterised by a set physiological responses. Although literature can sometimes provide contradictory information, a selection of 6 physiological measurements have been chosen in section 2.4 to characterise startle & surprise. For surprise these measures include: P300 ERP, increased theta and delta EEG frequency band activity and a specific HR pattern. For startle, these measures are: eye blinking, increased trapezius muscle activity, increased ratio of high frequency to low frequency HRV and an increased HR.

Since detecting S&S is theoretically possible by purely observing these physiological reactions, measuring these reactions can provide an objective reference for the development of the subjective S&S rating scale.

The effects can be measured using electromyography (EMG), electrocardiography (ECG) and electroencephalography (EEG). A research question was constructed, stating:

> **To which degree can surprise and startle be detected by designing and implementing a physiological sensor suite to perform EMG, ECG and EEG measurements?**

This report focused on selecting and theoretically developing a sensor suite that can be used to measure the aforementioned effects. Different options were compared against the desired general and technical requirement for each of the measurement types: ECG, EMG and EEG. The OpenBCI Cyton board was chosen due to its high accuracy, low price point and possibilities for integration into DUECA and custom development.

To answer the research question, 3 experiments will be performed in further research. The first experiment will serve to verify whether the data obtained from the Cyton board is of sufficient quality to extract HR, HRV, muscle activity and EEG frequency band activity. It will not involve stimuli of any sort.

Thereafter, the second experiment will use startling and surprising stimuli to verify whether the sensor suite is able to detect the chosen physiological responses connected to startle and surprise. A third experiment will serve as a validation experiment, where the same stimuli will be used to elicit S&S, but no information about the type of stimulus will be present during the data analysis phase. The goal of this experiment is to validate whether S&S can be detected reliably when nothing is known about the stimulus' nature.

Bibliography

Alberdi, A., Aztiria, A., & Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. Journal of Biomedical Informatics, 59, 49–75. https://doi.org/https://doi.org/10.1016/j.jbi.2015.11.007

Bailey, J. J., Berson, A. S., Garson, A., Horan, L. G., Macfarlane, P. W., Mortara, D. W., & Zywietz, C. (1990). Recommendations for standardization and specifications in automated electrocardiography: Bandwidth and digital signal processing. a report for health professionals by an ad hoc writing group of the committee on electrocardiography and cardiac electrophysiology of the council on clinical cardiology, american heart association. Circulation, 81(2), 730–739. https://doi.org/10.1161/01.CIR.81.2.730

Bainbridge, L. (1983). Ironies of automation. In G. JOHANNSEN & J. RIJNSDORP (Eds.), Analysis, design and evaluation of manmachine systems (pp. 129–135). Pergamon. https://doi.org/https://doi.org/10.1016/B978-0-08-029348-6.50026-9

Baar-Eroglu, C., Baar, E., Demiralp, T., & Schürmann, M. (1992). P300-response: Possible psychophysiological correlates in delta and theta frequency channels. a review. International Journal of Psychophysiology, 13(2), 161–179. https://doi.org/https://doi.org/10.1016/0167-8760(92)90055-G

Baar-Eroglu, C., Demiralp, T., Schürmann, M., & Baar, E. (2001). Topological distribution of oddball p300 responses. International Journal of Psychophysiology, 39(2), 213–220. https://doi.org/https://doi.org/10.1016/S0167-8760(00)00142-2

Batista, D., Plácido da Silva, H., Fred, A., Moreira, C., Reis, M., & Ferreira, H. (2019). Benchmarking of the bitalino biomedical toolkit against an established gold standard. 6, 32–36. https://doi.org/10.1049/htl.2018.5037

Batista, D., Silva, H., & Fred, A. (2017). Experimental characterization and analysis of the bitalino platforms against a reference device. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2418–2421. https://doi.org/10.1109/EMBC.2017.8037344

Belcastro, C., & Foster, J. (2010). Aircraft loss-of-control accident analysis. In Aiaa guidance, navigation, and control conference. https://doi.org/10.2514/6.2010-8004

Biopac Systems. (2023a). Bionomadix series product sheet. https://www.biopac.com/wp-content/uploads/BioNomadix-Series.pdf

Biopac Systems. (2023b). Bionomadix wireless ppg and eda amplifier. https://www.biopac.com/product/bionomadix-ppg-and-eda-amplifier/

Biopac Systems. (2023c). Data Acquisistion, Loggers, Amplifiers: Biopac. https://www.biopac.com/

Biopac Systems. (2023d). Mp160 data acquisition systems. https://www.biopac.com/product/mp150-data-acquisition-systems/

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of erp components a tutorial [Multivariate Decoding and Brain Reading]. NeuroImage, 56(2), 814–825. https://doi.org/https://doi.org/10.1016/j.neuroimage.2010.06.048

Blascovich, J., & Mendes, W. (2000). Feeling and thinking: The role of affect in social cognition. Studies in Emotion and Social Interaction, 59–82.

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. Psychophysiology, 42(1), 1–15. https://doi.org/https://doi.org/10.1111/j.1469-8986.2005.00271.x

Boeing Commercial Airplanes. (2020). STATISTICAL SUMMARY OF COMMERCIAL JET AIRPLANE ACCIDENTS: Worldwide Operations 1959 2020. Retrieved December 16, 2022, from https://skybrary.aero/sites/default/files/bookshelf/32664.pdf

Bogaerts, W. (2023). Brainflowtest/readcyton/cmakelists.txt.

Borst, C., Suijkerbuijk, H. C. H., Mulder, M., & Paassen, M. M. V. (2006). Ecological interface design for terrain awareness. The International Journal of Aviation Psychology, 16(4), 375–400. https://doi.org/10.1207/s15327108ijap1604_3

Boucsein, W. (2012). Methods of electrodermal recording. https://doi.org/10.1007/978-1-4614-1126-0_2

BrainFLow. (2023). Brainflow.

BrainFlow. (2019a). Code Samples: C++ Read Data from a Board. https://brainflow-openbci.readthedocs.io/en/latest/Examples.html#id2

BrainFlow. (2019b). Supported Boards. https://brainflow-openbci.readthedocs.io/en/latest/SupportedBoards.html#supported-boards

BrainFlow. (2023). BrainFlow, How Biosensors Work. https://brainflow.org/

Bruna, O., Levora, T., & Holub, J. (2018). Assessment of ecg and respiration recordings from simulated emergency landings of ultra light aircraft. Scientific Reports, 8. https://doi.org/10.1038/s41598-018-25528-z

Burki-Cohen, J. (2010). Technical challenges of upset recovery training: Simulating the element of surprise. In Aiaa modeling and simulation technologies conference. https://doi.org/10.2514/6.2010-8008

Burma, J. S., Lapointe, A. P., Soroush, A., Oni, I. K., Smirl, J. D., & Dunn, J. F. (2021). The validity and reliability of an open source biosensing board to quantify heart rate variability. Heliyon, 7(6), e07148. https://doi.org/https://doi.org/10.1016/j.heliyon.2021.e07148

Burns, A., Doheny, E., Greene, B., Foran, T., Leahy, D., O'Donovan, K., & Mcgrath, M. (2010). Shimmer: An extensible platform for physiological signal capture. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2010, 3759–62. https://doi.org/10.1109/IEMBS.2010.5627535

Camilo, E. M., Gutiérrez, J. A. M., Ramírez, O. P., Martínez, J. G., Hernández, A. V., & Salas, L. L. (2020). A functional electrical stimulation controller for contralateral hand movements based on emg signals. 2020 17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), 1–6. https://doi.org/10.1109/CCE50788.2020.9299199

Cao, J., Tian, Z., & Wang, Z. (2019). Hand gestures recognition based on one-channel surface emg signal*. Journal of Software Engineering and Applications, 12, 383–392. https://doi.org/10.4236/jsea.2019.129023

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. International Journal of Psychophysiology, 155, 49–62. https://doi.org/https://doi.org/10.1016/j.ijpsycho.2020.05.010

Cassani, R., Moinnereau, M.-A., Ivanescu, L., Rosanne, O., & Falk, T. H. (2020). Neural interface instrumented virtual reality headsets: Toward next-generation immersive applications. IEEE Systems, Man, and Cybernetics Magazine, 6(3), 20–28. https://doi.org/10.1109/MSMC.2019.2953627

Chen, Y., Yang, Z., & Yangliang, W. (2021). A soft exoskeleton glove for hand bilateral training via surface emg. Sensors, 21, 578. https://doi.org/10.3390/s21020578

Collet, C., Vernet-Maury, E., Delhomme, G., & Dittmar, A. (1997). Autonomic nervous system response patterns specifity to basic emotions. Journal of the Autonomic Nervous System, 62, 45–57. https://doi.org/10.1016/S0165-1838(96)00108-7

Diagnostic electrocardiographic devices. (1996).

Donchin, E., HEFFLEY, E., HILLYARD, S., LOVELESS, N., MALTZMAN, I., ÖHMAN, A., Roesler, F., RUCHKIN, D., & SIDDLE, D. (2006). Cognition and eventrelated potentials ii. the orienting reflex and p300. Annals of the New York Academy of Sciences, 425, 39–57. https://doi.org/10.1111/j.1749-6632.1984.tb23522.x

Dreissen, Y. E., Bakker, M. J., Koelman, J. H., & Tijssen, M. A. (2012). Exaggerated startle reactions. Clinical Neurophysiology, 123(1), 34–44. https://doi.org/https://doi.org/10.1016/j.clinph.2011.09.022

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, p300, and n400. Clinical Neurophysiology, 120(11), 1883–1908. https://doi.org/https://doi.org/10.1016/j.clinph.2009.07.045

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. Psychophysiology, 14(5), 456–467. https://doi.org/https://doi.org/10.1111/j.1469-8986.1977.tb01312.x

Emotiv. (2023a). Emotiv Epoc X. https://www.emotiv.com/epoc-x/

Emotiv. (2023b). Headset Comparison Chart. https://www.emotiv.com/comparison/

Epstein, S., Boudreau, L., & Kling, S. (1975). Magnitude of the heart rate and electrodermal response as a function of stimulus input, motor output, and their interaction. Psychophysiology, 12(1), 15–24. https://doi.org/https://doi.org/10.1111/j.1469-8986.1975.tb03053.x

European Aviation Safety Agency. (2015). Notice of Proposed Amendment 2015-13: Loss of control prevention and recovery training. Retrieved December 16, 2022, from https://www.easa.europa.eu/en/document-library/notices-of-proposed-amendment/npa-2015-13

Eysenck, M., Derakshan, N., Santos, R., & Calvo, M. (2007). Anxiety and cognitive performance: Attentional control theory. Emotion (Washington, D.C.), 7, 336–53. https://doi.org/10.1037/1528-3542.7.2.336

Federal Aviation Administration. (2015). AC 120-111 - Upset Prevention and Recovery Training - with Change. Retrieved December 16, 2022, from https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/1027328

Frey, J. (2016). Comparison of an open-hardware electroencephalography amplifier with medical grade device in brain-computer interface applications, 105–114. https://doi.org/10.5220/0005954501050114

Ghaderi, F., Kim, S. K., & Kirchner, E. A. (2014). Effects of eye artifact removal methods on single trial p300 detection, a comparative study. Journal of Neuroscience Methods, 221, 41–47. https://doi.org/https://doi.org/10.1016/j.jneumeth.2013.08.025

Gibson, Z., Butterfield, J., Rodger, M., Murphy, B., & Marzano, A. (2019). Use of dry electrode electroencephalography (eeg) to monitor pilot workload and distraction based on p300 responses to an auditory oddball task. In H. Ayaz & L. Mazur (Eds.), Advances in neuroergonomics and cognitive engineering (pp. 14–26). Springer International Publishing.

González-Mendoza, A., Pérez-SanPablo, A. I., López-Gutiérrez, R., & Quiñones-Urióstegui, I. (2018). Validation of an emg sensor for internet of things and robotics. 2018 15th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), 1–5. https://doi.org/10.1109/ICEEE.2018.8533972

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. Cognition and Emotion, 9(1), 87–108. https://doi.org/10.1080/02699939508408966

Guerrero, F. N., & Spinelli, E. M. (2022). Biopotential acquisition systems. In F. Simini & P. Bertemes-Filho (Eds.), Medicine-based informatics and engineering (pp. 51–79). Springer International Publishing. https://doi.org/10.1007/978-3-030-87845-0_4

Habibzadeh Tonekabony Shad, E., Molinas, M., & Ytterdal, T. (2020). Impedance and noise of passive and active dry eeg electrodes: A review. IEEE Sensors Journal, 20(24), 14565–14577. https://doi.org/10.1109/JSEN.2020.3012394

Hajcak, G., & Foti, D. (2008). Errors are aversive: Defensive motivation and the error-related negativity [PMID: 18271855]. Psychological Science, 19(2), 103–108. https://doi.org/10.1111/j.1467-9280.2008.02053.x

Halford, J. J., Sabau, D., Drislane, F. W., Tsuchida, T. N., & Sinha, S. R. (2016). American clinical neurophysiology society guideline 4: Recording clinical eeg on digital media [PMID: 28436799]. The Neurodiagnostic Journal, 56(4), 261–265. https://doi.org/10.1080/21646821.2016.1245563

Handbook of psychophysiology (4th ed.). (2016). Cambridge University Press. https://doi.org/10.1017/9781107415782

Hannula, M., Huttunen, K., Koskelo, J., Laitinen, T., & Leino, T. (2008). Comparison between artificial neural network and multilinear regression models in an evaluation of cognitive workload in a flight simulator. Computers in Biology and Medicine, 38(11), 1163–1170. https://doi.org/https://doi.org/10.1016/j.compbiomed.2008.09.007

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139–183). North-Holland. https://doi.org/https://doi.org/10.1016/S0166-4115(08)62386-9

Healey, J., & Picard, R. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems, 6(2), 156–166. https://doi.org/10.1109/TITS.2005.848368

Henckens, M., Hermans, E., Pu, Z., Joels, M., & Fernández, G. (2009). Stressed memories: How acute stress affects memory formation in humans. The Journal of neuroscience : the official journal of the Society for Neuroscience, 29, 10111–9. https://doi.org/10.1523/JNEUROSCI.1184-09.2009

Holand, S., Girard, A., Laude, D., Meyer-Bisch, C., & Elghozi, J.-L. (2000). Holand s, girard a, laude d, meyer-bisch c, elghozi jl. effects of an auditory startle stimulus on blood pressure and heart rate in humans. j hypertens 17: 1893-1897. Journal of hypertension, 17, 1893–7. https://doi.org/10.1097/00004872-199917060-00055

Horstmann, G., & Herwig, A. (2014). Surprise attracts the eyes and binds the gaze. Psychonomic bulletin & review, 22. https://doi.org/10.3758/s13423-014-0723-1

iMotions. (2020). Skin Conductance Response What it is and How to Measure it. Retrieved January 12, 2023, from https://imotions.com/blog/learning/best-practice/skin-conductance-response/

International Civil Aviation Organisation. (2013). Manual of evidence-based training (doc 9995). Version 1. ICAO.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention [Visual Attention: Psychophysics, electrophysiology and neuroimaging]. Vision Research, 49(10), 1295–1306. https://doi.org/https://doi.org/10.1016/j.visres.2008.09.007

Johnson, M. K., Blanco, J. A., Gentili, R. J., Jaquess, K. J., Oh, H., & Hatfield, B. D. (2015). Probe-independent eeg assessment of mental workload in pilots. 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), 581–584. https://doi.org/10.1109/NER.2015.7146689

Joshi, R. K., S, M. K., S, H. R., Jayachandra, M., & Pandya, H. J. (2022). Design, development and validation of a portable visual p300 event-related potential extraction system. 2022 IEEE Biomedical Circuits and Systems Conference (BioCAS), 409–413. https://doi.org/10.1109/BioCAS54905.2022.9948657

Juárez, S., González, A., Maya, M., Cardenas, A., Gonzalez-Galvan, E. J., & Medellin-Castillo, H. I. (2022). A comprehensive system for the acquisition of emg signals and muscle force in lower limb. 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 1–6. https://doi.org/10.1109/ICECCME55909.2022.9988114

Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. European Journal of Neuroscience, 41(8), 1068–1078. https://doi.org/https://doi.org/10.1111/ejn.12859

Koch, M. (1999). The neurobiology of startle. Progress in Neurobiology, 59(2), 107–128. https://doi.org/https://doi.org/10.1016/S0301-0082(98)00098-7

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review [The biopsychology of emotion: Current theoretical and empirical perspectives]. Biological Psychology, 84(3), 394–421. https://doi.org/https://doi.org/10.1016/j.biopsycho.2010.03.010

Kret, M., & Sjak-Shie, E. (2018). Preprocessing pupil size data: Guidelines and code. Behavior Research Methods, 51. https://doi.org/10.3758/s13428-018-1075-y

Kuchar, J., & Yang, L. (2000). A review of conflict detection and resolution modeling methods. IEEE Transactions on Intelligent Transportation Systems, 1(4), 179–189. https://doi.org/10.1109/6979.898217

Kuhn, M., Wendt, J., Sjouwerman, R., Büchel, C., Hamm, A., & Lonsdorf, T. B. (2020). The neurofunctional basis of affective startle modulation in humans: Evidence from combined facial electromyography and functional magnetic resonance imaging [Stress Mechanisms and Fear Memories]. Biological Psychiatry, 87(6), 548–558. https://doi.org/https://doi.org/10.1016/j.biopsych.2019.07.028

Ladd, C., Plotsky, P., & Davis, M. (2010). Startle response. Encyclopedia of stress, 3, 561–568. https://doi.org/10.1016/B978-012373947-6.00356-1

Lakhan, P., Banluesombatkul, N., Changniam, V., Dhithijaiyratn, R., Leelaarporn, P., Boonchieng, E., Hompoonsup, S., & Wilaiprasitporn, T. (2019). Consumer grade brain sensing for emotion recognition. IEEE Sensors Journal, 19(21), 9896–9907. https://doi.org/10.1109/JSEN.2019.2928781

Landman, A. (2019). Managing startle and surprise in the cockpit (Doctoral dissertation) [ISBN: 978-94-6182-963-4]. Delft University of Technology.

Landman, A., Groen, E. L., van Paassen, M. M. (, Bronkhorst, A. W., & Mulder, M. (2017a). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise [PMID: 28777917]. Human Factors, 59(8), 1161–1172. https://doi.org/10.1177/0018720817723428

Landman, A., Groen, E. L., van Paassen, M. M. (, Bronkhorst, A. W., & Mulder, M. (2017b). The influence of surprise on upset recovery performance in airline pilots. The Inter-

national Journal of Aerospace Psychology, 27(1-2), 2–14. https://doi.org/10.1080/10508414.2017.1365610

Landman, A., Oorschot, P., Van Paassen, M. M., Groen, E., Bronkhorst, A., & Mulder, M. (2018). Training pilots for unexpected events: A simulator study on the advantage of unpredictable and variable scenarios. Human Factors: The Journal of the Human Factors and Ergonomics Society, 60, 001872081877992. https://doi.org/10.1177/0018720818779928

Landman, A., van Middelaar, S. H., Groen, E. L., van Paassen, M. M. (, Bronkhorst, A. W., & Mulder, M. (2020). The effectiveness of a mnemonic-type startle and surprise management procedure for pilots. The International Journal of Aerospace Psychology, 30(3-4), 104–118. https://doi.org/10.1080/24721840.2020.1763798

Leuchs, L., Schneider, M., & Spoormaker, V. I. (2019). Measuring the conditioned response: A comparison of pupillometry, skin conductance, and startle electromyography. Psychophysiology, 56(1), e13283. https://doi.org/https://doi.org/10.1111/psyp.13283

Linden, D. E. J. (2005). The p300: Where in the brain is it produced and what does it tell us? [PMID: 16282597]. The Neuroscientist, 11(6), 563–576. https://doi.org/10.1177/1073858405280524

Ma'as, M. D. F., Masitoh, Azmi, A. Z. U., & Suprijanto. (2017). Real-time muscle fatigue monitoring based on median frequency of electromyography signal. 2017 5th International Conference on Instrumentation, Control, and Automation (ICA), 135–139. https://doi.org/10.1109/ICA.2017.8068428

Malik, M., & Camm, A. J. (1990). Heart rate variability. Clinical Cardiology, 13(8), 570–576. https://doi.org/https://doi.org/10.1002/clc.4960130811

Man, P.-K., Cheung, K.-L., Sangsiri, N., Shek, W. J., Wong, K.-L., Chin, J.-W., Chan, T.-T., & So, R. H.-Y. (2022). Blood pressure measurement: From cuff-based to contactless monitoring. Healthcare, 10(10), 2113. https://doi.org/10.3390/healthcare10102113

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of eye-related measures of drivers mental workload [6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015]. Procedia Manufacturing, 3, 2854–2861. https://doi.org/https://doi.org/10.1016/j.promfg.2015.07.783

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. Y. (2016). A flight simulator study of the impairment effects of startle on pilots during unexpected critical events. Aviation Psychology and Applied Human Factors, 6(1), 24–32. https://doi.org/10.1027/2192-0923/a000092

Medical electrical equipment - part 2-25: Particular requirements for the basic safety and essential performance of electrocardiographs. (2011).

Medical electrical equipment part 1-11: General requirements for basic safety and essential performance collateral standard: Requirements for medical electrical equipment and medical electrical systems used in the home healthcare environment. (2015).

Meyer, W.-U., Niepel, M., Rudolph, U., & Schützwohl, A. (1991). An experimental analysis of surprise. Cognition and Emotion, 5(4), 295–311. https://doi.org/10.1080/02699939108411042

Mitello, L. G., & Hutton, R. J. B. (1998). Applied cognitive task analysis (acta): A practitioner's toolkit for understanding cognitive task demands [PMID: 9819578]. Ergonomics, 41(11), 1618–1641. https://doi.org/10.1080/001401398186108

Mohd Azli, M. A. S., Mustafa, M., Abdubrani, R., Abdul Hadi, A., Syed Ahmad, S. N. A., & Zahari, Z. L. (2019). Electromyograph (emg) signal analysis to predict muscle fatigue during driving. In Z. Md Zain, H. Ahmad, D. Pebrianti, M. Mustafa, N. R. H. Abdullah,

R. Samad, & M. Mat Noh (Eds.), Proceedings of the 10th national technical seminar on underwater system technology 2018 (pp. 405–420). Springer Singapore.

Mulert, C., Pogarell, O., Juckel, G., Rujescu, D., Giegling, I., Rupp, D., Mavrogiorgou, P., Bussfeld, P., Gallinat, J., Möller, H., & Hegerl, U. (2004). The neural basis of the p300 potential: Focus on the time-course of the underlying cortical generators. European archives of psychiatry and clinical neuroscience, 254, 190–8. https://doi.org/10.1007/s00406-004-0469-2

MyoWare. (2022). MyoWare 2.0 Advanced Guide. https://cdn.sparkfun.com/assets/learn_tutorials/1/9/5/6/MyoWare_v2_AdvancedGuide-Updated.pdf

Nmcová, A., Maránová, L., Smisek, R., Vitek, M., & Koláová, J. (2016). Recommendations for ecg acquisition using bitalino. Proceedings of the 22nd Conference STUDENT EEICT, 543–547.

Niso, G., Romero, E., Moreau, J. T., Araujo, A., & Krol, L. R. (2022). Wireless eeg: An survey of systems and studies. NeuroImage, 119774. https://doi.org/https://doi.org/10.1016/j.neuroimage.2022.119774

Noordewier, M. K., Topolinski, S., & Van Dijk, E. (2016). The temporal dynamics of surprise. Social and Personality Psychology Compass, 10(3), 136–149. https://doi.org/https://doi.org/10.1111/spc3.12242

Ogedegbe, G., & Pickering, T. (2010). Principles and techniques of blood pressure measurement. Cardiology clinics, 28, 571–86. https://doi.org/10.1016/j.ccl.2010.07.006

Olofsson, F. (2019). OpenBCI WiFi Shield DIY. https://fredrikolofsson.com/f0blog/openbci-wifi-shield-diy/

OpenBCI. (2022a). Cyton Data Format. https://docs.openbci.com/Cyton/CytonDataFormat/

OpenBCI. (2022b). OpenBCI Documentation: Cyton Board. https://docs.openbci.com/Cyton/CytonLanding/

OpenBCI. (2022c). OpenBCI Documentation: WiFi Shield Getting Started Guide. https://docs.openbci.com/GettingStarted/Boards/WiFiGS/

OpenBCI. (2023a). Cyton Getting Started Guide. https://docs.openbci.com/GettingStarted/Boards/CytonGS/

OpenBCI. (2023b). Installing the OpenBCI GUI as a "Standalone" Application. https://docs.openbci.com/Software/OpenBCISoftware/GUIDocs/#installing-the-openbci-gui-as-a-standalone-application

OpenBCI. (2023c). Linux Users: Serial Port Permissions. https://docs.openbci.com/Software/OpenBCISoftware/GUIDocs/#linux-users-serial-port-permissions

OpenBCI. (2023d). OpenBCI Cyton Biosensing Board (8-channels). https://shop.openbci.com/products/cyton-biosensing-board-8-channel

OpenBCI. (2023e). OpenBCI Ganglion Board (4-channels). https://shop.openbci.com/products/ganglion-board

Peterson, V., Galván, C., Hernández, H., & Spies, R. (2020). A feasibility study of a complete low-cost consumer-grade brain-computer interface system. Heliyon, 6(3), e03425. https://doi.org/https://doi.org/10.1016/j.heliyon.2020.e03425

Picton, T. (1992). The p300 wave of the human event-related potential. Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society, 9, 456–79. https://doi.org/10.1097/00004691-199210000-00002

Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R., Miller, G., Ritter, W., Ruchkin, D., Rugg, M., & Taylor, M. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. Psychophysiology, 37, 127–52. https://doi.org/10.1017/S0048577200000305

Plácido da Silva, H., Guerreiro, J., Lourenco, A., Fred, A., & Martins, R. (2014). Bitalino: A novel hardware framework for physiological computing. PhyCS 2014 - Proceedings of the International Conference on Physiological Computing Systems.

Plux Biosignals. (2016a). BITalino (r)evolution Plugged Kit Data Sheet. https://support.pluxbiosignals.com/wp-content/uploads/2021/11/revolution-bitalino-plugged-kit-datasheet.pdf

Plux Biosignals. (2016b). Electromyography (EMG) Sensor Data Sheet. https://support.pluxbiosignals.com/wp-content/uploads/2021/11/revolution-emg-sensor-datasheet-1.pdf

Plux Biosignals. (2018). Electroencephalography (EEG) Sensor Data Sheet. https://support.pluxbiosignals.com/wp-content/uploads/2021/11/revolution-eeg-sensor-datasheet-revb.pdf

Plux Biosignals. (2020). Electrocardiography (ECG) Sensor Data Sheet. https://support.pluxbiosignals.com/wp-content/uploads/2021/11/revolution-ecg-sensor-datasheet-revb-1.pdf

Plux Biosignals. (2023). BITalino (r)evolution Board Kit BLE/BT. https://www.pluxbiosignals.com/products/bitalino-revolution-board-kit-ble-bt

pluxbiosignals. (2019). cpp-samples (Github repository). https://github.com/pluxbiosignals/cpp-samples

Preuschoff, K., t Hart, B., & Einhauser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenalines role in decision making. Frontiers in Neuroscience, 5. https://doi.org/10.3389/fnins.2011.00115

Rashid, U., Niazi, I., Signal, N., & Taylor, D. (2018). An eeg experimental study evaluating the performance of texas instruments ads1299. Sensors, 18(11), 3721. https://doi.org/10.3390/s18113721

Reisenzein, R., Bördgen, S., Holtbernd, T., & Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: The case of surprise. Journal of personality and social psychology, 91, 295–315. https://doi.org/10.1037/0022-3514.91.2.295

Reisenzein, R., Horstmann, G., & Schützwohl, A. (2019). The cognitive-evolutionary model of surprise: A review of the evidence. Topics in Cognitive Science, 11(1), 50–74. https://doi.org/https://doi.org/10.1111/tops.12292

Rinold, T., & Porges, S. W. (1997). Inferential and descriptive influences on measures of respiratory sinus arrhythmia: Sampling rate, r-wave trigger accuracy, and variance estimates. Psychophysiology, 34(5), 613–621. https://doi.org/https://doi.org/10.1111/j.1469-8986.1997.tb01748.x

Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F., & Yeh, M. (2014). Startle and surprise on the flight deck: Similarities, differences, and prevalence. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58(1), 1047–1051. https://doi.org/10.1177/1541931214581219

Rojas, G., Alvarez, C., Montoya Moya, C., de la Iglesia Vaya, M., Cisternas, J., & Gálvez, M. (2018). Study of resting-state functional connectivity networks using eeg electrodes position as seed. Frontiers in Neuroscience, 12. https://doi.org/10.3389/fnins.2018.00235

Rooseleer, F., Kirwan, B., Humm, E., & Alarcon, D. (2022). 'the application of human factors in wake vortex encounter flight simulations for the reduction of flight upset risk and startle response. https://doi.org/10.54941/ahfe1001565

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. Inter-

national Journal of Industrial Ergonomics, 35(11), 991–1009. https://doi.org/https://doi.org/10.1016/j.ergon.2005.04.005

Sänger, J., Bechtold, L., Schoofs, D., Blaszkewicz, M., & Wascher, E. (2014). The influence of acute stress on attention mechanisms and its electrophysiological correlates. Frontiers in Behavioral Neuroscience, 8. https://doi.org/10.3389/fnbeh.2014.00353

Shimmer Sensing. (2017a). Emg user guide. https://shimmersensing.com/wp-content/docs/support/documentation/EMG_User_Guide_Rev1.12.pdf

Shimmer Sensing. (2017b). Logandstream for shimmer3 firmware user manual. https://shimmersensing.com/wp-content/docs/support/documentation/LogAndStream_for_Shimmer3_Firmware_User_Manual_rev0.11a.pdf

Shimmer Sensing. (2017c). Shimmer user manual. https://bmslab.utwente.nl/wp-content/uploads/2019/12/Shimmer-User-manual.pdf

Shimmer Sensing. (2018). Ecg user guide. https://shimmersensing.com/wp-content/docs/support/documentation/ECG_User_Guide_Rev1.12.pdf

Shimmer Sensing. (2022a). Shimmer3 ecg unit. https://shimmersensing.com/product/shimmer3-ecg-unit-2/

Shimmer Sensing. (2022b). Shimmer3 emg unit. https://shimmersensing.com/product/shimmer3-emg-unit/

Siedlecka, E., & Denson, T. (2018). Experimental methods for inducing basic emotions: A qualitative review. Emotion Review, 11, 175407391774901. https://doi.org/10.1177/1754073917749016

Simons, R. F., Graham, F. K., Miles, M. A., & Balaban, M. T. (1998). Input and central processing expressed in erp and heart rate changes to rare target and rare nontarget stimuli. Psychophysiology, 35(5), 563–575. https://doi.org/https://doi.org/10.1017/S0048577298971566

Stegeman, D., & Hermens, H. (2007). Standards for suface electromyography: The european project surface emg for non-invasive assessment of muscles (seniam). 1.

stephanie E. Wemm, E. W. (2017). Effects of acute stress on decision making. Appl Psychophysiol Biofeedback, 42(1), 1–12. https://doi.org/10.1007/s10484-016-9347-8

Tajdini, M., Sokolov, V., Kuzminykh, I., Shiaeles, S., & Ghita, B. (2020). Wireless sensors for brain activity-a survey. Electronics, 9. https://doi.org/10.3390/electronics9122092

Texas Instruments. (2017). ADS1299-x Low-Noise, 4-, 6-, 8-Channel, 24-Bit, Analog-to-Digital Converter for EEG and Biopotential Measurements. https://www.ti.com/lit/ds/symlink/ads1299.pdf

Tichon, J., Mavin, T., Wallis, G., Visser, T., & Riek, S. (2014). Using pupillometry and electromyography to track positive and negative affect during flight simulation. Aviation Psychology and Applied Human Factors, 4, 23. https://doi.org/10.1027/2192-0923/a000052

Topolinski, S., & Strack, F. (2015). Corrugator activity confirms immediate negative affect in surprise. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.00134

Ugarte, D. E., Linares, D., Kemper, G., & Almenara, C. A. (2019). An algorithm to measure the stress level from eeg, emg and hrv signals. 2019 International Conference on Information Systems and Computer Science (INCISCOS), 346–353. https://doi.org/10.1109/INCISCOS49368.2019.00061

Valderrama, M., González, E., & García, F. (2021). Development of a low-cost surface emg acquisition system device for wearable applications. 2021 IEEE 2nd International Congress of Biomedical Engineering and Bioengineering (CI-IB&BI), 1–4. https://doi.org/10.1109/CI-IBBI54220.2021.9626100

Veltman, J. A. (2002). A comparative study of psychophysiological reactions during simulator and real flight. The International Journal of Aviation Psychology, 12(1), 33–48. https://doi.org/10.1207/S15327108IJAP1201_4

Vicente, K., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. IEEE Transactions on Systems, Man, and Cybernetics, 22(4), 589–606. https://doi.org/10.1109/21.156574

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. Ergonomics, 23(10), 995–1011. https://doi.org/10.1080/00140138008924809

Wijsman, J., Grundlehner, B., Penders, J., & Hermens, H. (2010). Trapezius muscle emg as predictor of mental stress. ACM Transactions on Embedded Computing Systems (TECS), 12, 155–163. https://doi.org/10.1145/2485984.2485987

Willems, J. L. (1985). Common standards for auantitative electrocardiography. Journal of Medical Engineering & Technology, 9(5), 209–217. https://doi.org/10.3109/03091908509032101

Wolf, H., Graystone, P., LeBlanc, A., Lywood, D., Raber, M., & Sherwood, J. (1976). Suggested minimum performance characteristics of data acquisition instrumentation in computer-assisted ecg processing systems. Journal of Electrocardiology, 9(3), 239–247. https://doi.org/https://doi.org/10.1016/S0022-0736(76)80051-9

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance [PMID: 12502155]. Human Factors, 44(3), 365–375. https://doi.org/10.1518/0018720024497709

Zhou, W., & Gotman, J. (2009). Automatic removal of eye movement artifacts from the eeg using ica and the dipole model. Progress in Natural Science, 19(9), 1165–1170. https://doi.org/https://doi.org/10.1016/j.pnsc.2008.11.013

References

[1] Frances K. Graham. “Constraints on Measuring Heart Rate and Period Sequentially Through Real and Cardiac Time”. In: *Psychophysiology* 15.5 (1978), pp. 492–495. DOI: https://doi.org/10.1111/j.1469-8986.1978.tb01422.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.1978.tb01422.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8986.1978.tb01422.x.

Part III

Appendices

A

Participant information letter

PARTICIPANT INFORMATION LETTER

Detecting startle & surprise using EEG, EMG and ECG

Hello,

You are being invited to participate in a research study titled "detecting startle & surprise using EEG, EMG and ECG". The experiment will be performed by Ward Bogaerts (TU Delft MSc student), under supervision of René van Paassen and Olaf Stroosma. Information regarding the experiment is provided in this letter. If you have any questions about this experiment, don't hesitate to contact the researcher.

Background of the research

Understanding the influence of startle and surprise (S&S) allows to construct more effective pilot training programs, thereby reducing the risk of fatal accidents caused by unexpected events. These events induce both cognitive and physical reactions. A subjective rating scale to indicate startle and surprise, similar to the NASA TLX scale for measuring workload, is currently being developed. A set of physiological measurements can assist in developing this scale by introducing objective quantitative data in addition to subjective data.

Purpose of the research

The main goal of this experiment is validating a physiological sensor suite to measure the physiological effects of startle and surprise and assist in developing the subjective scale. The data will be utilized for scientific studies and/or publications. The researcher will never publish any information which can be traced back to you in any report or publication.

Risks of participating

The goal of the experiment is to detect startle and surprise. You will be subjected to startling events which can increase stress or anxiety levels in some cases. The experiment is not set up to explicitly induce these effects but nevertheless you might experience the initial symptoms such as increased sweating, heart rate and breathing rate. If you suffer from either epilepsy, hyperekplexia, colour-blindness or any heart disease, you are not eligible to participate. The experiment will be performed using battery-powered experimental equipment not certified for consumer use. In the unlikely case of coming in contact with the power source, an electrical shock may be felt. However, voltages are low and should not cause any permanent harm. Electrodes will be placed on the body, which can cause mild discomfort and leave a small gel residue when removed. A small clamp will be attached to the earlobe, so removal of earrings might be necessary. The set-up will involve cables suspended from the body, which imposes a small risk of tripping when standing up. Participation is voluntary, and you have the right to revoke it at any time.

COVID-19 transmission is a possibility as a result of the (still ongoing) pandemic of the coronavirus. The Control and Simulations section at Aerospace Engineering has established COVID safety guidelines, and these will be followed throughout the experiment in order to reduce this risk. Ventilation will be done in the appropriate places before entry.

In case of an emergency, the TU Delft BHV staff will be contacted on the number 88888 to provide timely assistance.

What does participation in the research involve?

The experiment will take place in the SIMONA building at the Faculty of Aerospace Engineering of TU Delft. You will be sitting behind a desk, facing a display. No movement or control actions are required to conduct the experiment. Electrodes will be attached underneath the right eye, on the chest and hip, on the back shoulder muscle and on the scalp.

You will perform a counting task. A sequence of coloured dots will be presented on the display. Green dots can be ignored, while the number of appearances of the red dot should be counted. Only a single, or no dot, will be shown at any given time. The location of the dots on the screen will not change. Counting has to be performed mentally, it should not be said out loud. At the end of the experiment, you will be asked to report the number of times a red dot appeared. In case of any discomfort, a short break may be requested verbally. The experiment consists of 1 session and will take approximately 20 minutes.

Procedures for withdrawal from the study

Your participation in this study is fully voluntary, and you are free to end it at any time, including in the middle of the experiment, by verbally indicating this to the researcher. You have the right to ask for personal data access, correction, or deletion. You are not required to provide justification for your choice. To do this, get in touch with the researchers using the details provided at the end of this document.

Confidentiality of data

It is required to gather and use the following personal information for this investigation: name, e-mail address, age, gender, and whether you suffer from epilepsy, hyperplexia, colour-blindness or any heart disease. We shall take the necessary security precautions to protect your personal information and ensure its confidentiality. This means that your data will be kept in a safe storage environment at TU Delft at all times. Only the researchers will have access to the data. All information will be handled in confidence and kept in a participant-only database. Only on the informed consent form will your name be connected to a participant number. The informed consent form will be kept in a separate, secure location and kept digitally. Your information will remain private in this manner. The only people that know your participant number are the researchers.

The personal data will be retained for linking your participation number to the informed consent, to facilitate the erasure of personal details, if you request.

The findings of this investigation will possibly be reported in upcoming scientific journals. Any publications (master's thesis report, scientific papers, reports) about the study will never include your participant number or name.

Contact Information

If you have any complaints regarding confidentiality of your data, you can contact the Aerospace faculty's data steward, Heather Andrews, via (H.E.AndrewsMancilla@tudelft.nl).

Researcher's names, numbers and email addresses:

Ward Bogaerts:

w.bogaerts@student.tudelft.nl

[redacted]

René van Paassen:

M.M.vanPaassen@tudelft.nl

Olaf Stroosma:

O.Stroosma@tudelft.nl

B

Informed consent form

.

Consent Form for Detecting startle & surprise using EEG, EMG and ECG

Researchers:	**Ward Bogaerts**
Title of research:	*Detecting startle & surprise using EEG, EMG and ECG*
Supervisors:	**René van Paassen , Olaf Stroosma**

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
Risks associated with participating in the study		
I understand that taking part in this study involves the following risks: increased levels of stress or anxiety, mild discomfort	○	○
I understand that taking part in the study involves the very small risk of electrocution if coming in contact with the measurement equipment power source	○	○
I understand that the measurements can be invasive and cause mild discomfort	○	○
Use of the information in the study		
I understand that the data that I provide will be used for scientific reports and or publications and that the researcher will not identify me in any report or publication that will result from this experiment and that my confidentiality as a participant in this study remains secure	○	○
I understand that personal information collected about me that can identify me, will not be shared beyond the study team.	○	○
Future use and reuse of the information by others		
I give permission for the anonymized data that I provide to be archived so it can be used for future research and learning.	○	○
Safety		
I confirm that the researcher has provided me with detailed safety briefing and operational instructions to guarantee that the experiment can be performed in line with the current TU Delft COVID-19 guidelines and that I have understood these instructions, and that this experiment shall at all times follow the TU Delft guidelines	○	○

I understand that this research study has been reviewed and approved by the TU Delft Human Research Ethics Committee (HREC). I am aware that I can report any problems regarding my participation in the experiment to the researchers using the contact information below. ○ ○

Signatures

_____	_____	_____
Name of participant:	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Researcher name:	Signature	Date

Study contact details for further information:

Ward Bogaerts

[redacted]

w.bogaerts@student.tudelft.nl

René van Paassen

M.M.vanPaassen@tudelft.nl

Olaf Stroosma

O.Stroosma@tudelft.nl

C

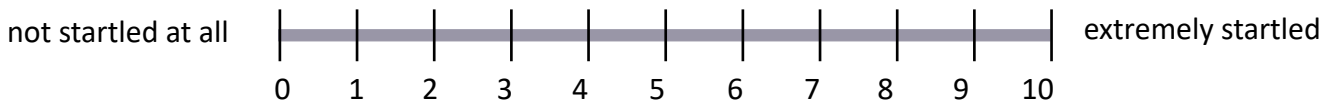
Post-experiment questionnaire

Startle & Surprise questionnaire

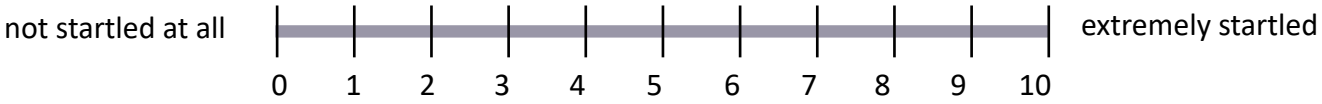
You can mark your answer anywhere on the scales below. You do not have to use the ticks.

Startle

Please indicate, by marking a point on the line below, how **startled** you were by the **first** bang.

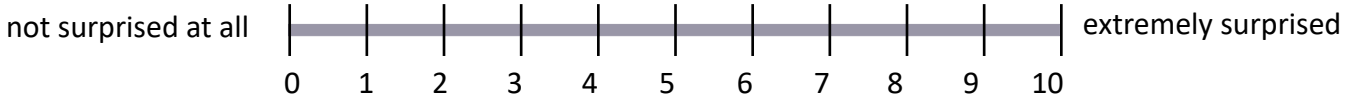


Please indicate, by marking a point on the line below, how **startled** you were by the **last** bang.

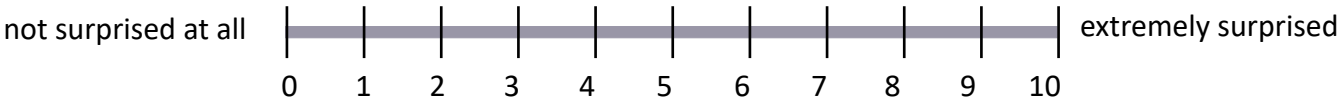


Surprise

Please indicate, by marking a point on the line below, how **surprised** you were by the **first** blue dot.



Please indicate, by marking a point on the line below, how **surprised** you were by the **last** blue dot.



[to be filled in by researcher]

Participant number:

Startle sequence:

Surprise sequence:

D

OpenBCI Cyton device report

Delft University of Technology
INSPECTION REPORT FOR DEVICES TO BE USED IN CONNECTION WITH HUMAN SUBJECT RESEARCH

This report should be completed for every experimental device that is to be used in interaction with humans and that is not CE certified or used in a setting where the CE certification no longer applies[1].

The first part of the report has to be completed by the researcher and/or a responsible technician.

Then, the safety officer (Heath, Security and Environment advisor) of the faculty responsible for the device has to inspect the device and fill in the second part of this form. An actual list of safety-officers is provided on this webpage.

Note that in addition to this, all experiments that involve human subjects have to be approved by the Human Research Ethics Committee of TU Delft. Information on ethics topics, including the application process, is provided on the HREC website.

Device identification (name, location): OpenBCI Cyton,

Configurations inspected[2]: 4 EEG leads on Fz, Pz, Cz and O& locations. 2 ECG leads, 2 EMG leads. 1 BIAS lead and 1 SRB2 lead

Type of experiment to be carried out on the device:[3] Sitting in front of a digital screen

Name(s) of applicants(s): Ward Bogaerts, René van Paassen

Job title(s) of applicants(s): MSc student, associate professor

(Please note that the inspection report should be filled in by a TU Delft employee. In case of a BSc/MSc thesis project, the responsible supervisor has to fill in and sign the inspection report.)

Date:

Signature(s):

---

1 Modified, altered, used for a purpose not reasonably foreseen in the CE certification

2 If the devices can be used in multiple configurations, otherwise insert NA

3 e.g. driving, flying, VR navigation, physical exercise, ...

Setup summary

Please provide a brief description of the experimental device (functions and components) and the setup in which context it supposed to be used. Please document with pictures where necessary.

More elaborate descriptions should be added as an appendix (see below).

An OpenBCI Cyton board (figures 1 and 2) will be used to obtain 4 EEG, 1 ECG and 2 EMG measurements. It is battery operated using a 3.7V 1S LiPo. The board and battery are contained in a plastic cover to reduce the risk of touching electrically live components. Communication happens via BLE to a USB dongle, connected to a PC (figure 5). Electrode cables are attached to the I/O pins at the bottom of the board. All electrode cables use a quick snap system to attach electrodes. EEG electrodes consist of 4 dry electrodes mounted in a head band, together with 2 dry reference electrode clamps to be placed on the earlobes. The electrodes are passive and only measure potentials. Figure 3 shows the 4 EEG electrodes together with the reference electrode clamps. 4 electrodes for 2 different measurement locations will be used for EMG: two small electrodes to be placed under the eye and 2 larger electrodes to be placed on the trapezius muscle. 2 ECG electrodes are used, to be placed on the right chest and left hip. The EMG and ECG electrodes can be seen in figure 4. The participant will be sitting behind a desk with the electrodes attached. Medical tape can be used to prevent the electrode cables from swaying and to keep them in place. The participant will in no case have to handle the OpenBCI Cyton board. The board will only be operated by the researchers.

The experiments will be conducted from May 2023 until October 2023 and therefore it is requested to inspect the device for this duration.

In case of emergency, the TU Delft BHV staff will be contacted by dialing 88888 in order to provide timely assistance.

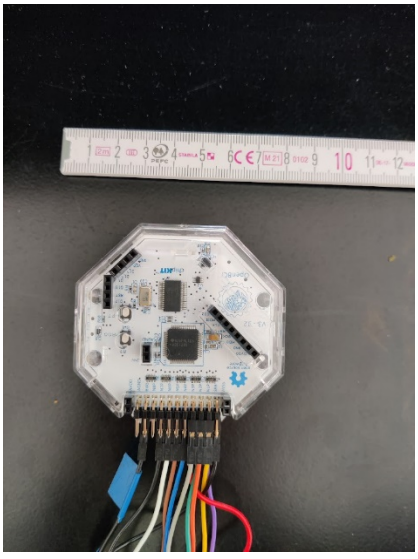


Figure 1: OpenBCI Cyton front view, in plastic casing with electrode wires attached

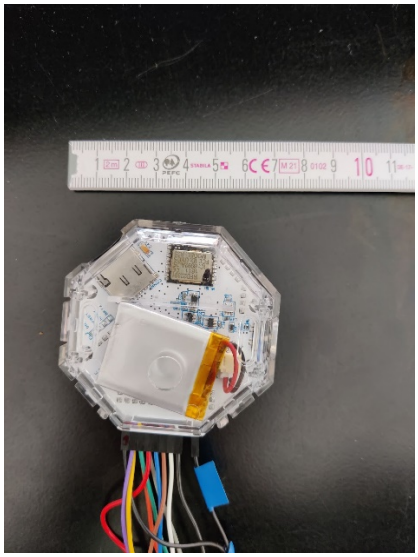


Figure 2: OpenBCI Cyton in plastic casing (rear view)

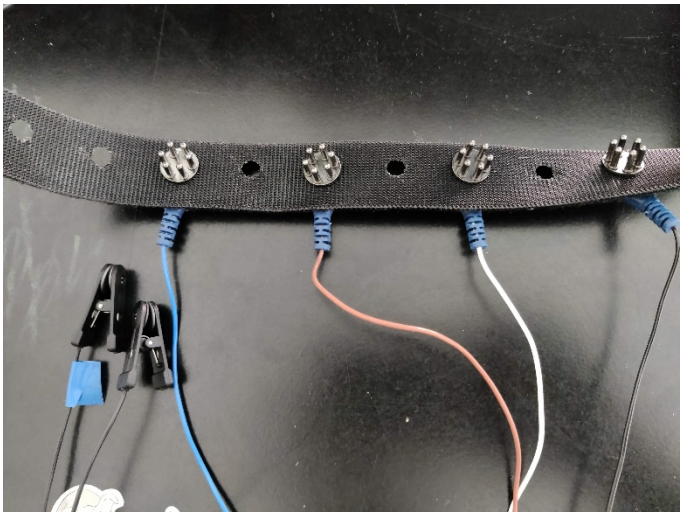


Figure 3: 4 dry EEG electrodes placed in a headband (top) and 2 dry reference clamp electrodes (bottom)

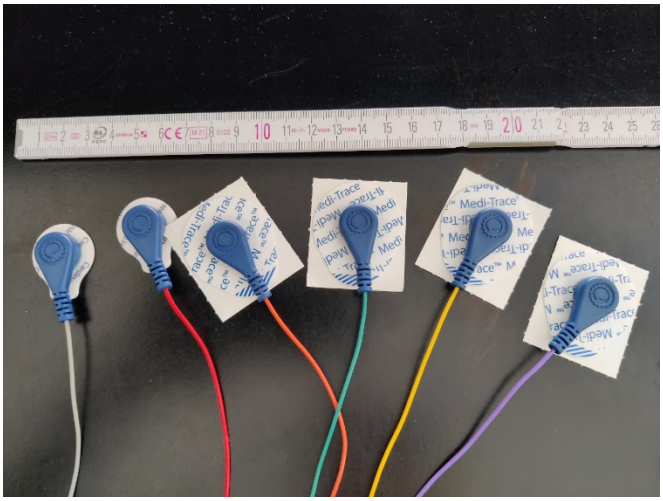


Figure 4: EMG eye blink electrodes (grey, red), EMG trapezius electrodes (orange, green) and ECG electrodes (yellow, purple)

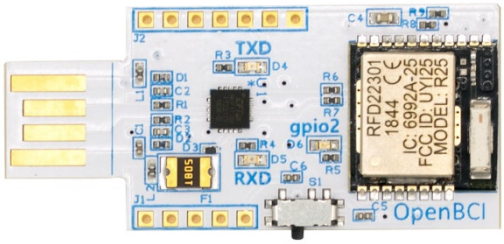


Figure 5: BLE dongle (to be connected to a PC)

Risk checklist

Please fill in the following checklist and consider these hazards that are typically present in many research setups. If a hazard is present, please describe how it is dealt with.

Also, mention any other hazards that are present.

Hazard type	Present	Hazard source	Mitigation measures
Mechanical (sharp edges, moving equipment, etc.)	No		
Electrical	Yes	OpenBCI Cyton battery	Enclosed in plastic casing. Only handled by researchers. Low voltage
Structural failure	No		
Touch Temperature	Yes	OpenBCI Cyton microchip	Enclosed in plastic casing
Electromagnetic radiation	No		
Ionizing radiation	No		
(Near-)optical radiation (lasers, IR-, UV-, bright visible light sources)	No		
Noise exposure	No		
Materials (flammability, offgassing, etc.)	Yes	OpenBCI Cyton battery	Balanced charging, keeping voltages at correct levels and not fully depleting the battery. Visual inspection whether battery is swollen before the start of the experiment.
Chemical processes	No		
Fall risk	No		
Other:			
Other:			

Appendices

The OpenBCI Cyton operates on 3.7V, which is below the DC safe handling voltage of the EN-50110-1 norm.

Electrodes will first be attached to the corresponding electrode cables before they are placed on the body.

New EMG/ECG electrodes will be used for every participant. The EEG electrodes will be cleaned with a soap-water solution.

Before the experiment, the battery will be charged using a dedicated LiPo charger.

Participants will be asked to remove earrings for the experiment, such that they do not interfere with the reference electrodes.

Participants will not come in direct contact with the OpenBCI Cyton board, only with the electrodes and cables.

Device inspection

(to be filled in by the AMA advisor of the corresponding faculty)

Name: Erwin van Rijn (HSE adviseur)

Faculty: AE

The device and its surroundings described above have been inspected. During this inspection I could not detect any extraordinary risks.

(Briefly describe what components have been inspected and to what extent (i.e. visually, mechanical testing, measurements for electrical safety etc.)

Date: 16 May 2023

Signature:  **Erwin van rijn**

Inspection valid until[4]: November 1, 2023

Note: changes to the device or set-up, or use of the device for an experiment type that it was not inspected for require a renewed inspection

4 Indicate validity of the inspection, with a maximum of 3 years

E

Human Research Ethics Committee approval form

Date 07-Jul-2023
Contact person Grace van Arkel, Policy Advisor Academic Integrity
E-mail E.G.vanArkel@tudelft.nl

Human Research Ethics Committee TU Delft
(http://hrec.tudelft.nl)

Visiting address
Jaffalaan 5 (building 31)
2628 BX Delft

Postal address
P.O. Box 5015 2600 GA Delft
The Netherlands

Ethics Approval Application: Detecting startle & surprise using EEG, EMG and ECG
Applicant: Stroosma, Olaf

Dear Olaf Stroosma,

It is a pleasure to inform you that your application mentioned above has been approved.

Thanks very much for your submission to the HREC which has been approved.

In addition to any specific conditions or notes, the HREC provides the following standard advice to all applicants:
• In light of recent tax changes, we advise that you confirm any proposed remuneration of research subjects with your faculty contract manager before going ahead.
• Please make sure when you carry out your research that you confirm contemporary covid protocols with your faculty HSE advisor, and that ongoing covid risks and precautions are flagged in the informed consent - with particular attention to this where there are physically vulnerable (eg: elderly or with underlying conditions) participants involved.
• Our default advice is not to publish transcripts or transcript summaries, but to retain these privately for specific purposes/checking; and if they are to be made public then only if fully anonymised and the transcript/summary itself approved by participants for specific purpose.
• Where there are collaborating (including funding) partners, appropriate formal agreements including clarity on responsibilities, including data ownership, responsibilities and access, should be in place and that relevant aspects of such agreements (such as access to raw or other data) are clear in the Informed Consent.

Good luck with your research!

Sincerely,

Dr. Ir. U. Pesch
Chair HREC
Faculty of Technology, Policy and Management

F

Cyton connections and channel set-up

Table F.1 summarises which channels on the OpenBCI Cyton board correspond to which measurment location. In addition, channel settings are summarised for each channel. These settings wee sent to the board in the ReadCyton Delft University Environment for Communication and Activation (DUECA) module. More information on channel settings can be found in the OenBCI Cyton Board SDK.

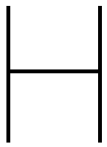
Channel number	Measurement	Gain	BIAS	SRB1	SRB2
1	ECG	24	Yes	No	No
2	EMG trapezius	24	Yes	No	No
3	EMG eye blink	24	Yes	No	No
4	EEG Fz	6	Yes	No	Yes
5	EEG Pz	6	Yes	No	Yes
6	EEG Cz	6	Yes	No	Yes
7	EEG O1	24	Yes	No	Yes
SRB2	Right mastoid (small gel electrode)	N.A.	N.A	N.A	N.A
BIAS	Left earlobe (dry clip electrode)	N.A.	N.A	N.A	N.A
SRB1	Not connected	N.A.	N.A	N.A	N.A

Table F.1: OpenBCI Cyton channel settings

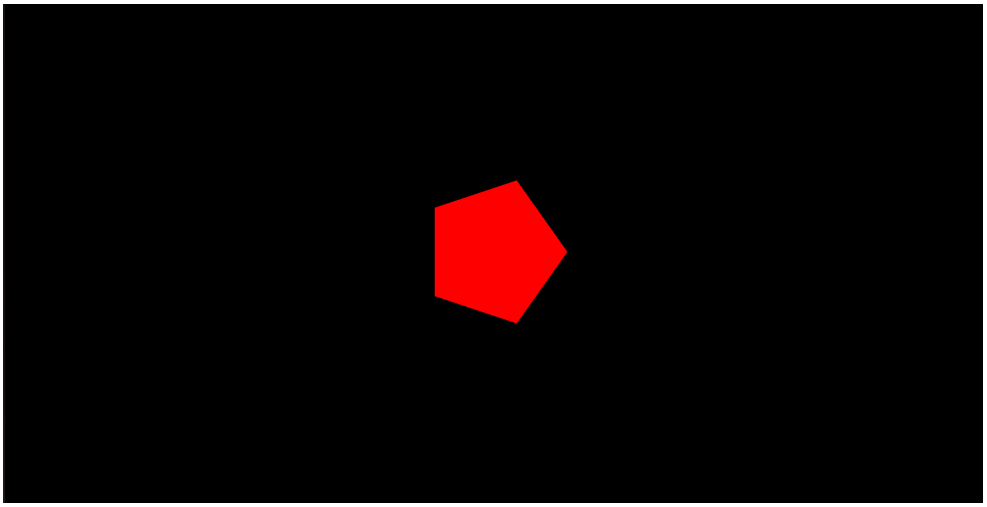
G

Experiment set-up

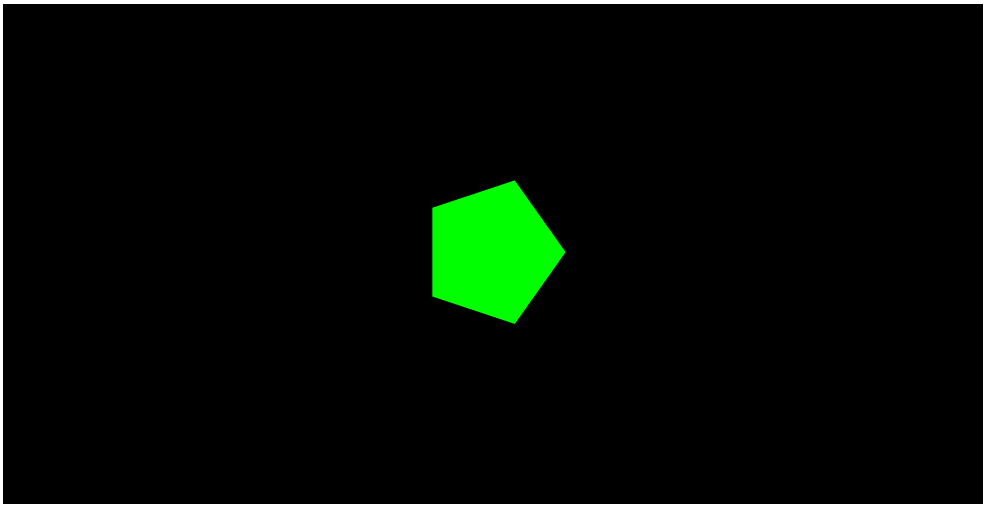
Figure G.1: Experiment set-up. Stimuli were presented on the computer display and via the stereo 5.0 system in a darkened room.



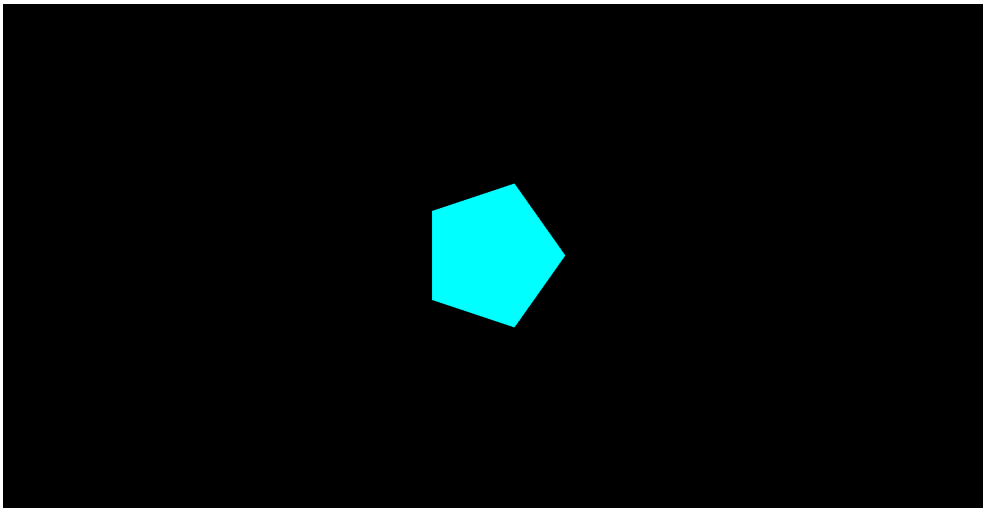
Surprise stimuli



(a) Frequent target stimulus



(b) Frequent non-target stimulus



(c) Rare non-target stimulus

Figure H.1: Surprise stimuli as presented to the participant

Data pre-processing

This chapter provides the results of signal filtering described in Section **??** of the scientific article. Figures are shown for participant 10, but results are comparable across participants.

I.1. EMG pre-processing

Figure I.1 shows the results of filtering and amplitude calculation for the EMG orbucilaris oculi channel, used for eye blink detection. The raw data from Figure I.1a shows a lot of drift artefacts, while the filtered signal is centred around 0. Figure I.1b show the effect of calculating the amplitude using the linear envelope, six clear spikes in the signal can already be observed. The smooth peaks seen in Figure I.1c can then be used to compare pre-stimulus amplitude with post-stimulus amplitude. The same results hold for the EMG trapezius channel data, shown in Figure I.2.

I.2. ECG pre-processing

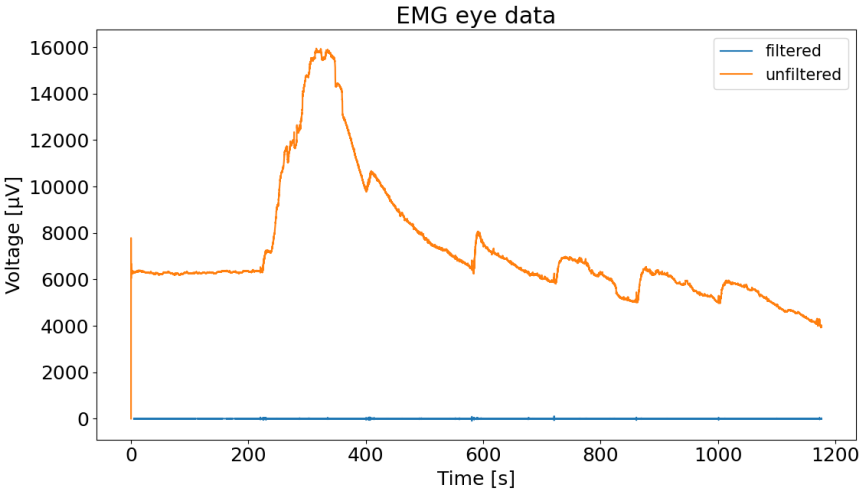
The pre-processing results for the ECG channel can be found in Figure I.3. Figures I.3a and I.3b show the raw and filtered ECG data. In Figure I.3c, the result of calculating the instant HR using 200 ms R-peaks epochs based on the method from [1] can be seen.

I.3. EEG pre-processing

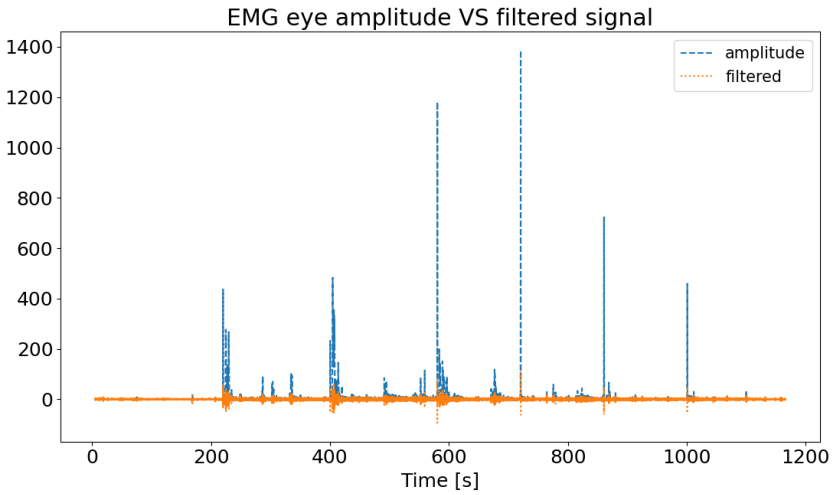
EEG data were pre-processed for multiple uses:

1. EEG data to be used along with independent component analysis (ICA) (Fz, Cz and Pz electrodes)
2. EEG data to be used to detect an increase in delta (1-3 Hz) and theta (3-6 Hz) frequency band activity (Fz, Cz and Pz electrodes)
3. EEG data to be used to detect eye blinks (O1 electrode)

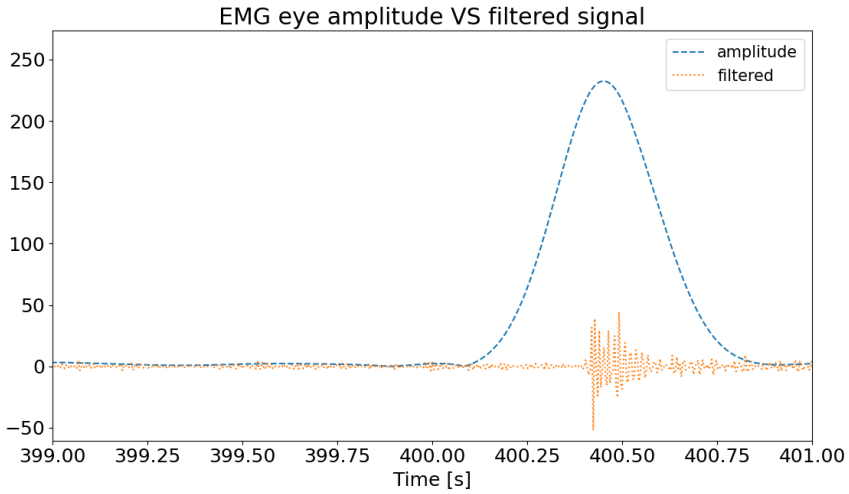
Figures I.4, I.5 and I.6 show the raw and filtered EEG data from the Fz, Cz and Pz electrodes. This filtered data was used later on in the ICA decomposition to detect the P300 event-related potential (ERP). Figure I.7 shows the EEG Fz, Cz and PZ data after filtering in the delta frequency band (1-3 Hz). Similarly, Figure I.8 shows the EEG Fz, Cz and Pz data after filtering in the theta frequency band (3-6Hz). Figure I.9 shows the raw and filtered EEG O1 data used for detecting eye blinks. Figures I.9b and I.9c show the amplitude obtained by calculating the linear envelope, in the same way as for EMG data.



(a) Raw and filtered EMG orbicularis oculi data

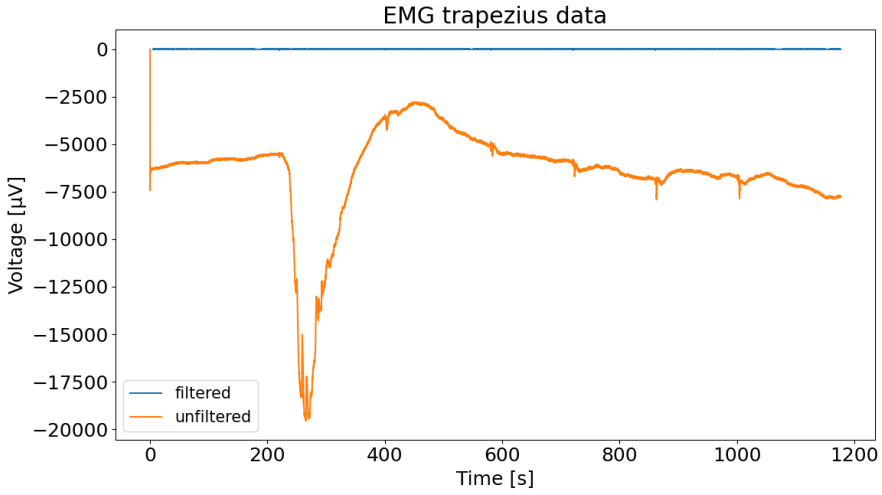


(b) Filtered EMG orbicularis oculi data VS its amplitude

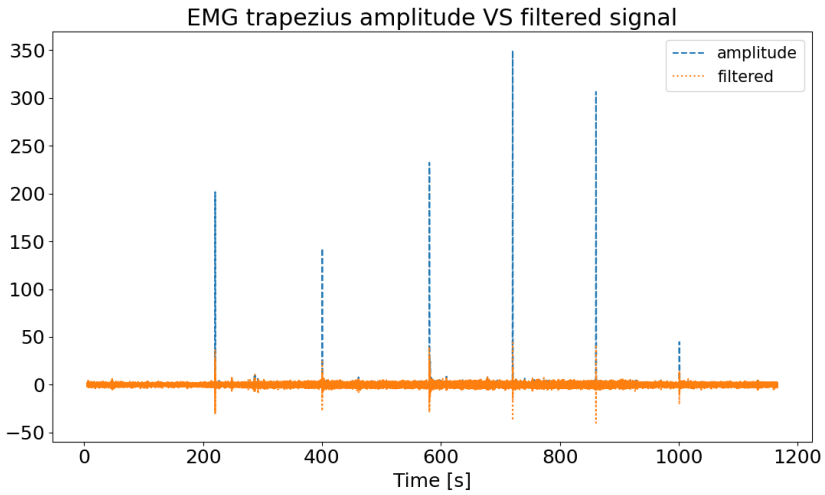


(c) Filtered EMG orbicularis oculi data VS its amplitude (zoomed)

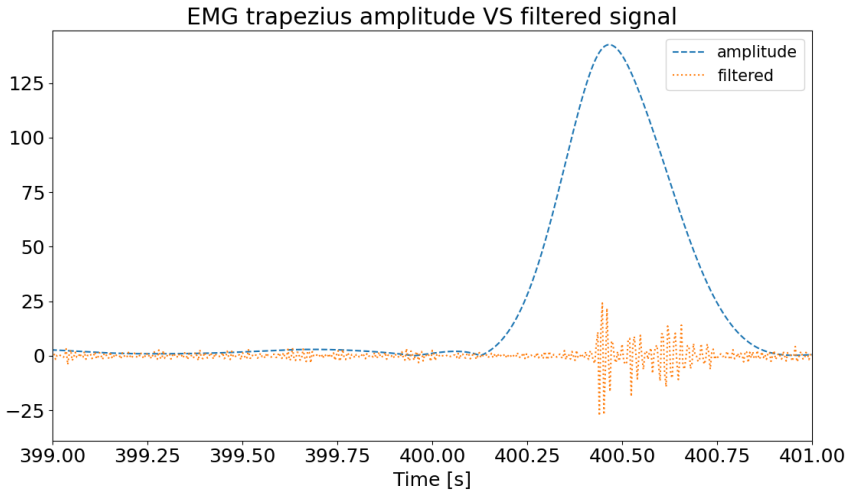
Figure I.1: Pre-processing of EMG orbicularis oculi data



(a) Raw and filtered EMG trapezius data

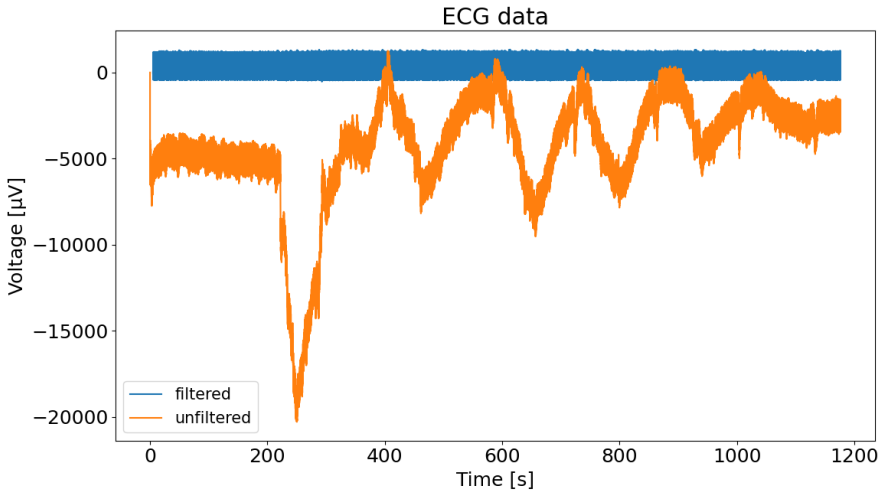


(b) Filtered EMG trapezius data VS its amplitude

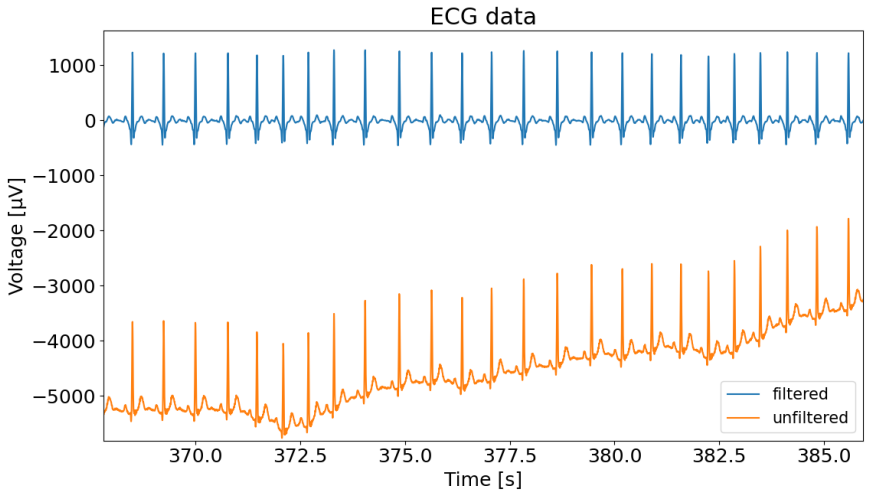


(c) Filtered EMG trapezius data VS its amplitude (zoomed)

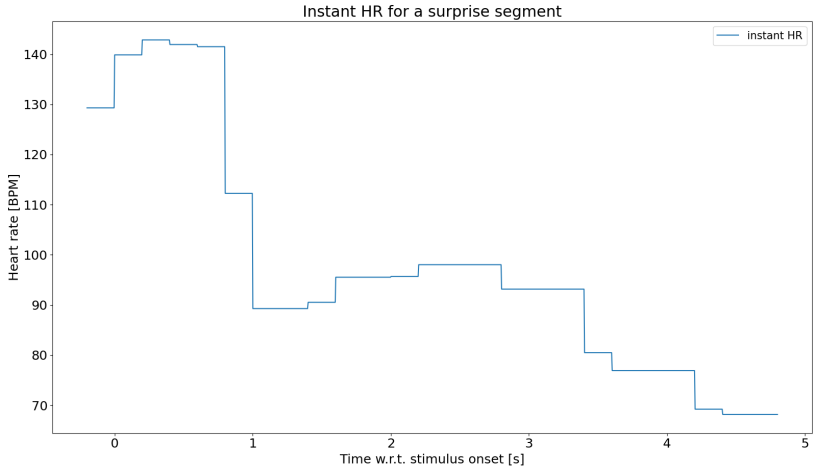
Figure I.2: Pre-processing of EMG trapezius data



(a) Raw and filtered ECG data

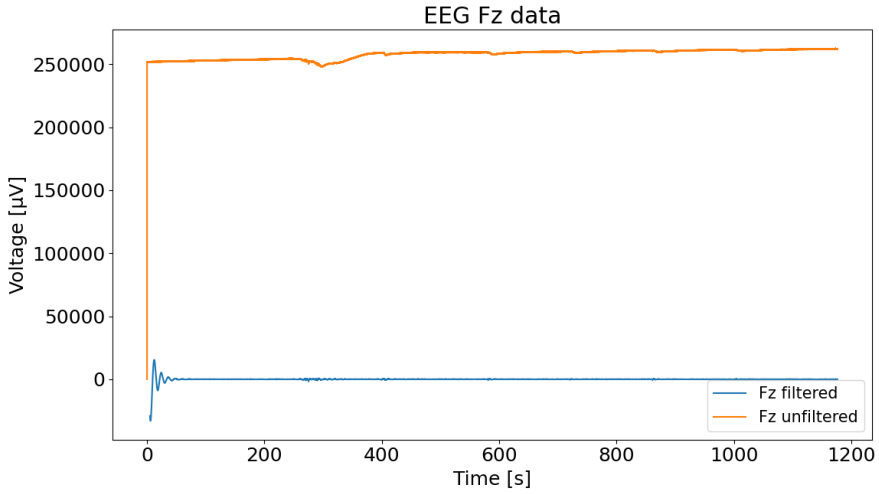


(b) Raw and filtered ECG data (zoomed)

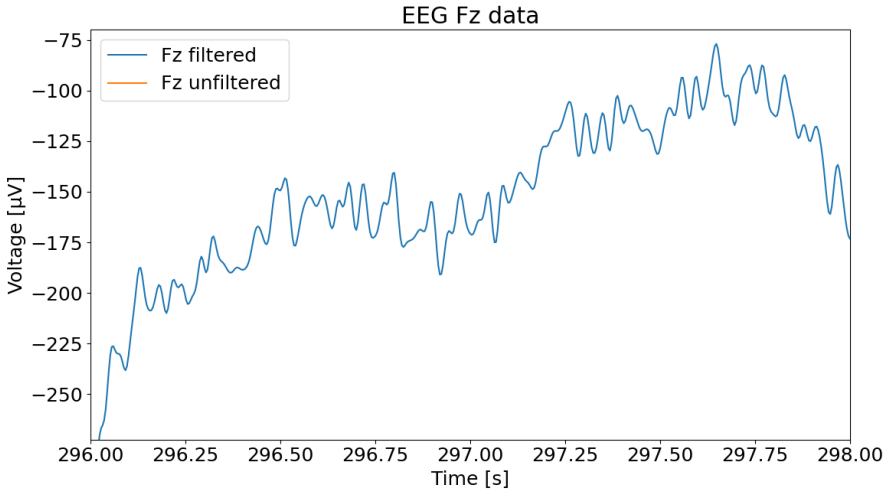


(c) Result of the instant HR calculation using the method from [1] for a single surprise segment

Figure I.3: Pre-processing of ECG data

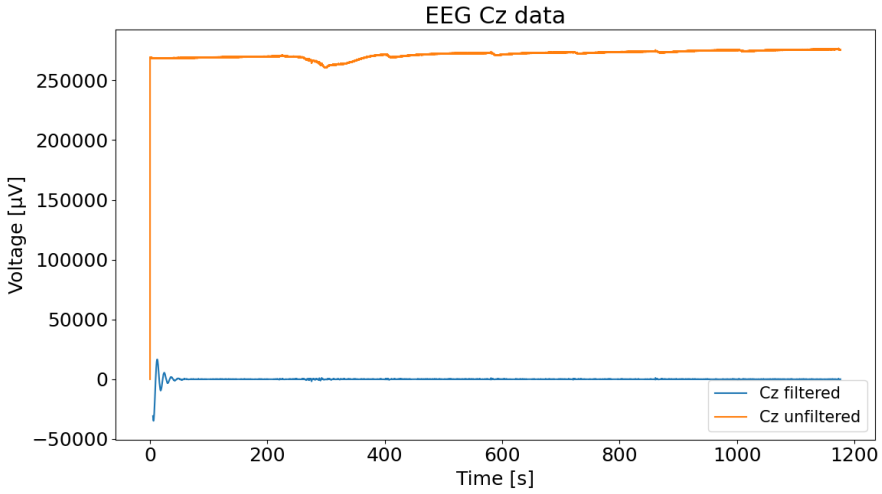


(a) Raw and filtered EEG Fz data

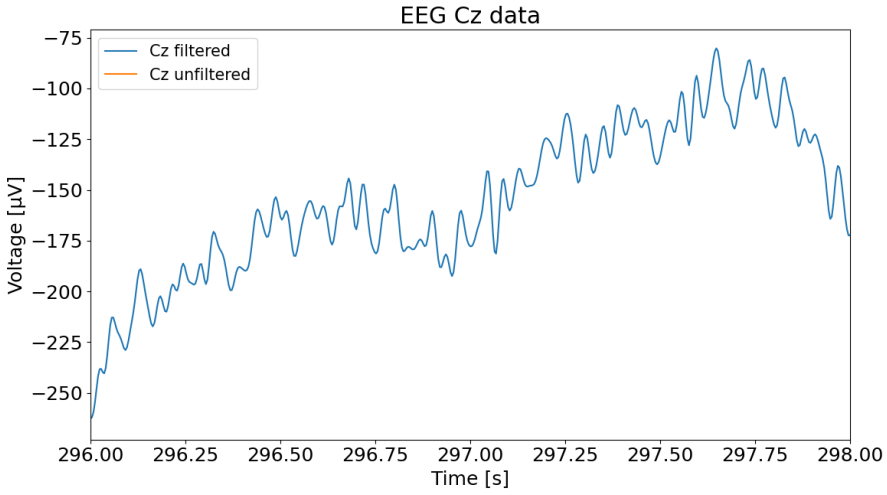


(b) Raw and filtered EEG Fz data (zoomed)

Figure I.4: Pre-processing of EEG Fz data to be used with ICA

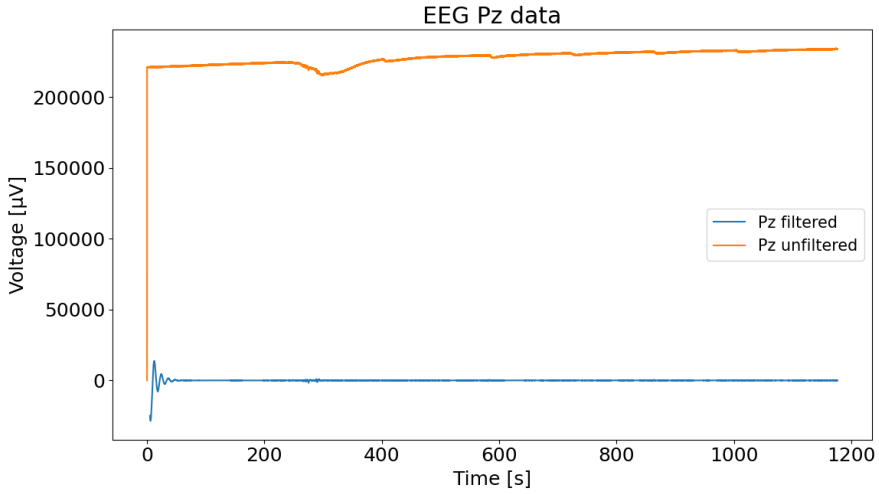


(a) Raw and filtered EEG Cz data

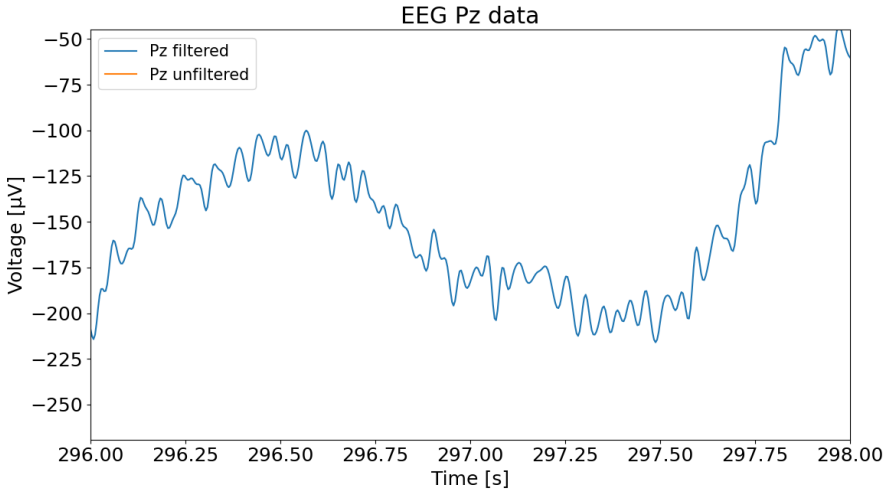


(b) Raw and filtered EEG Cz data (zoomed)

Figure I.5: Pre-processing of EEG Cz data to be used with ICA

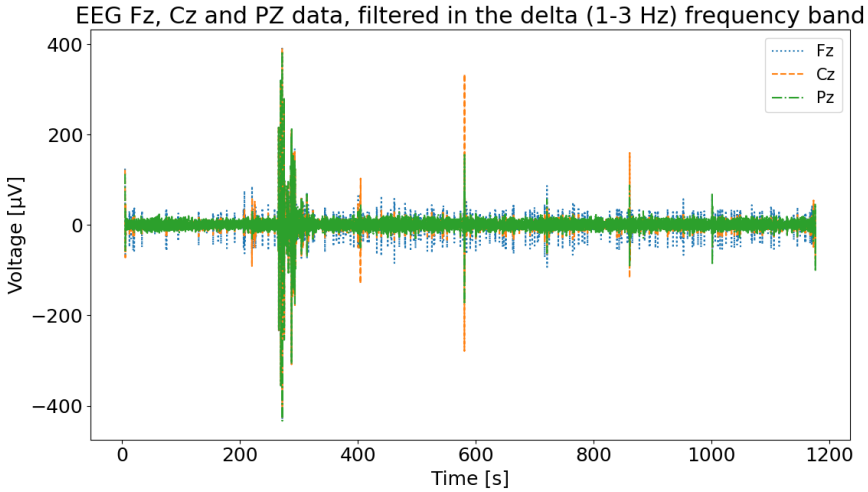


(a) Raw and filtered EEG Pz data

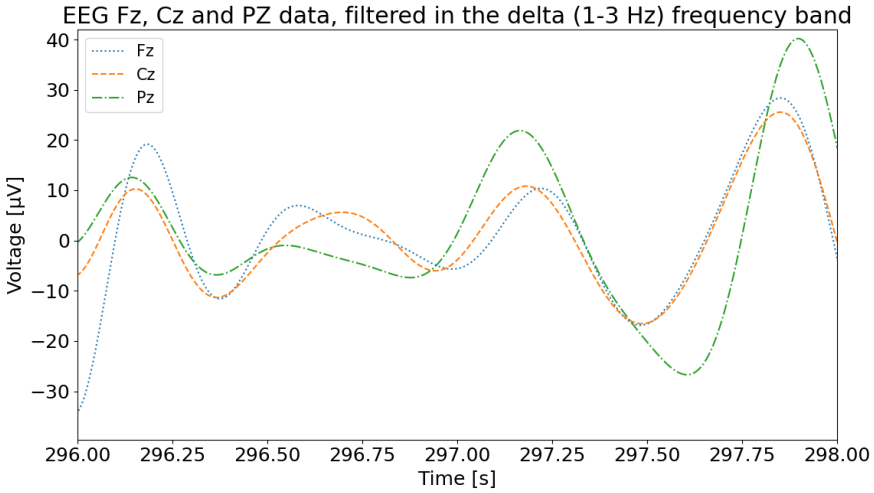


(b) Raw and filtered EEG Pz data (zoomed)

Figure I.6: Pre-processing of EEG Pz data to be used with ICA

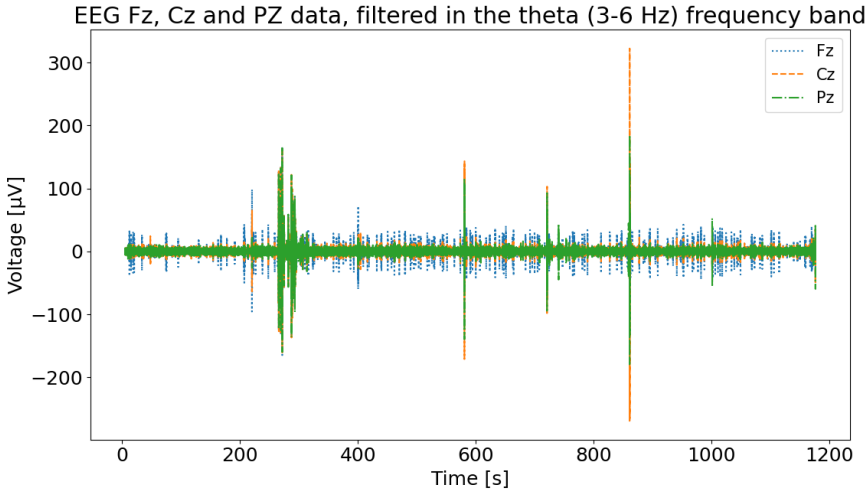


(a) Raw and filtered EEG Fz, Cz and PZ data in the delta (1-3 Hz) frequency band

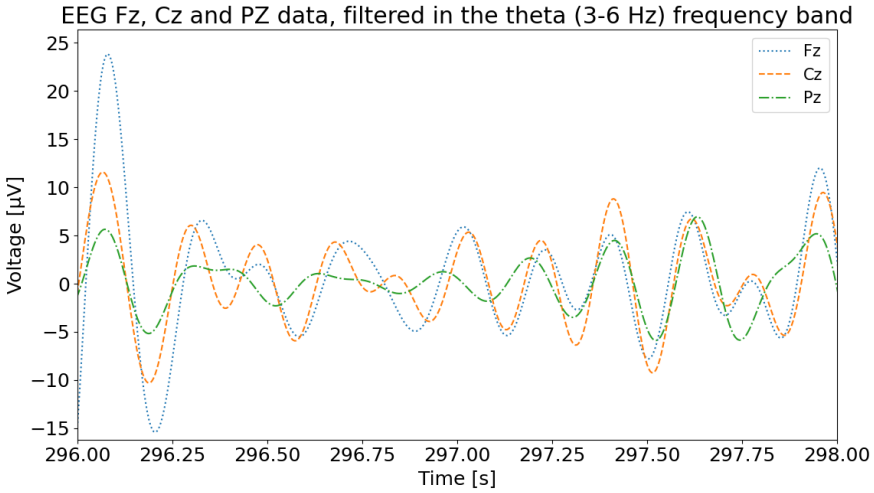


(b) Raw and filtered EEG Fz, Cz and PZ data in the delta (1-3 Hz) frequency band (zoomed)

Figure I.7: Pre-processing of EEG Fz, CZ and PZ data in the delta (1-3 Hz) frequency band

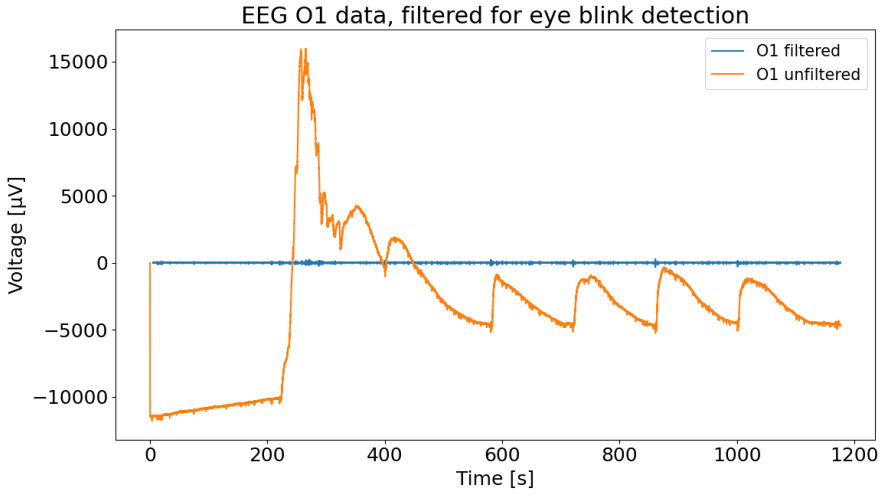


(a) Raw and filtered EEG Fz, Cz and PZ data in the theta (3-6 Hz) frequency band

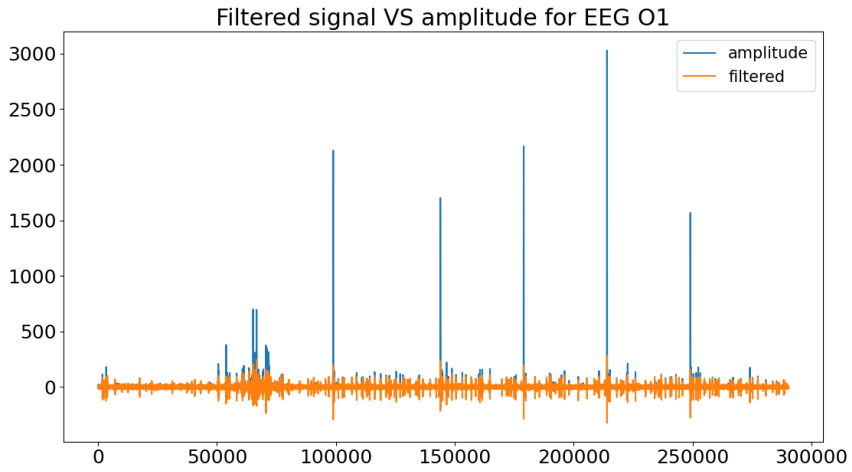


(b) Raw and filtered EEG Fz, Cz and PZ data in the theta (3-6 Hz) frequency band (zoomed)

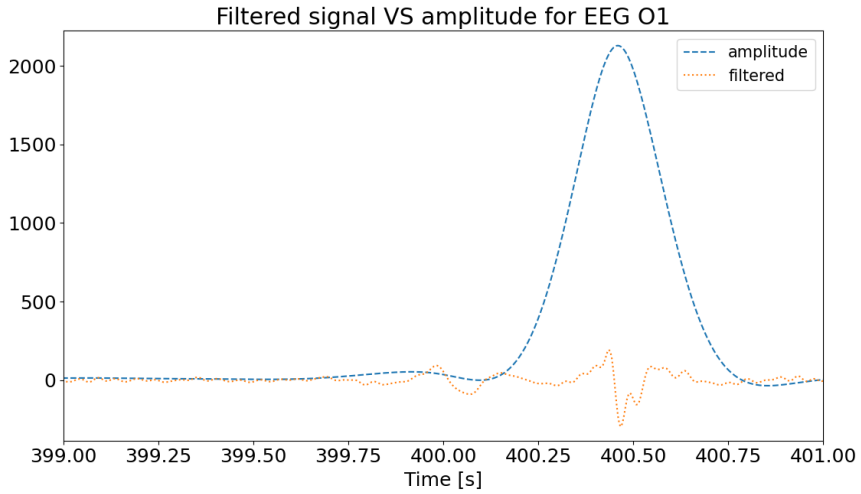
Figure I.8: Pre-processing of EEG Fz, CZ and PZ data in the theta (3-6 Hz) frequency band



(a) Raw and filtered EEG O1 data



(b) Filtered EEG O1 data VS its amplitude



(c) Filtered EEG O1 data VS its amplitude (zoomed)

Figure I.9: Pre-processing of EEG O1 data

J

Results of normality tests

This chapter provides the results of the Shapiro-Wilk normality tests for all data sequences considered in the validation of the physiological effects (Section **??** in the scientific article). For all tests, the significance threshold was set to $\alpha = 0.05$; if $p > \alpha$, the hypothesis indicating that the data is not normally distributed was rejected.

J.1. Startle validation

Table J.1: Results of pre-and post-stimulus data normality tests using the Shapiro-Wilk test for startle effects

	p	W-statistic
EMG eye amplitude pre-stimulus	< 0.01	0.697
EMG eye amplitude post-stimulus	< 0.01	0.644
EMG trapezius amplitude pre-stimulus	< 0.01	0.340
EMG trapezius amplitude post-stimulus	< 0.01	0.359
HR pre-stimulus	0.844	0.976
HR post-stimulus	0.042	0.908
Questionnaire response first stimulus	0.063	0.919
Questionnaire response last stimulus	0.090	0.926

J.2. Surprise validation

Table J.2: Results of pre-and post-stimulus data normality tests using the Shapiro-Wilk test for surprise effects. Results for rare non-target stimuli

	p	W-statistic
Average PTP delta band activity pre-stimulus	0.020	0.892
Average PTP delta band activity post-stimulus	< 0.01	0.847
First stimulus PTP delta band activity pre-stimulus	< 0.01	0.734
First stimulus PTP delta band activity post-stimulus	< 0.01	0.630
Average PTP theta band activity pre-stimulus	< 0.01	0.403
Average PTP theta band activity post-stimulus	< 0.01	0.376
First stimulus PTP theta band activity pre-stimulus	< 0.01	0.855
First stimulus PTP theta band activity post-stimulus	< 0.01	0.837
Average INT delta band activity pre-stimulus	0.021	0.892
Average INT delta band activity post-stimulus	0.220	0.942
First stimulus INT delta band activity pre-stimulus	< 0.01	0.691
First stimulus INT delta band activity post-stimulus	< 0.01	0.694
Average INT theta band activity pre-stimulus	< 0.01	0.548
Average INT theta band activity post-stimulus	< 0.01	0.382
First stimulus INT theta band activity pre-stimulus	0.112	0.928
First stimulus INT theta band activity post-stimulus	< 0.01	0.823
ΔHR for decrease	< 0.01	0.816
ΔHR for increase	0.792	0.974
Questionnaire response first stimulus	0.627	0.967
Questionnaire response last stimulus	< 0.01	0.738

Table J.3: Results of pre-and post-stimulus data normality tests using the Shapiro-Wilk test for surprise effects. Results for frequent non-target stimuli

	p	W-statistic
Average PTP delta band activity pre-stimulus	< 0.01	0.846
Average PTP delta band activity post-stimulus	< 0.01	0.799
First stimulus PTP delta band activity pre-stimulus	0.042	0.907
First stimulus PTP delta band activity post-stimulus	< 0.01	0.855
Average PTP theta band activity pre-stimulus	< 0.01	0.875
Average PTP theta band activity post-stimulus	< 0.01	0.834
First stimulus PTP theta band activity pre-stimulus	< 0.01	0.814
First stimulus PTP theta band activity post-stimulus	< 0.01	0.842
Average INT delta band activity pre-stimulus	0.023	0.895
Average INT delta band activity post-stimulus	< 0.01	0.862
First stimulus INT delta band activity pre-stimulus	0.160	0.936
First stimulus INT delta band activity post-stimulus	0.745	0.971
Average INT theta band activity pre-stimulus	0.195	0.939
Average INT theta band activity post-stimulus	0.018	0.889
First stimulus INT theta band activity pre-stimulus	0.027	0.898
First stimulus INT theta band activity post-stimulus	0.092	0.924
ΔHR for decrease	0.587	0.964
ΔHR for increase	0.143	0.933

Table J.4: Results of pre-and post-stimulus data normality tests using the Shapiro-Wilk test for surprise effects. Results for frequent target stimuli

	p	W-statistic
Average PTP delta band activity pre-stimulus	0.215	0.942
Average PTP delta band activity post-stimulus	0.051	0.911
First stimulus PTP delta band activity pre-stimulus	0.023	0.894
First stimulus PTP delta band activity post-stimulus	< 0.01	0.825
Average PTP theta band activity pre-stimulus	0.084	0.922
Average PTP theta band activity post-stimulus	< 0.01	0.868
First stimulus PTP theta band activity pre-stimulus	< 0.01	0.872
First stimulus PTP theta band activity post-stimulus	< 0.01	0.864
Average INT delta band activity pre-stimulus	0.131	0.931
Average INT delta band activity post-stimulus	0.251	0.945
First stimulus INT delta band activity pre-stimulus	< 0.01	0.870
First stimulus INT delta band activity post-stimulus	0.513	0.961
Average INT theta band activity pre-stimulus	0.429	0.957
Average INT theta band activity post-stimulus	0.032	0.901
First stimulus INT theta band activity pre-stimulus	0.018	0.889
First stimulus INT theta band activity post-stimulus	0.031	0.901
ΔHR for decrease	0.113	0.928
ΔHR for increase	0.091	0.924

K

Cost function output

The number of correct and false detections in function of the cost function parameter X can be seen in Figure K.1 for startle and Figure K.2 for surprise. The correct and false detections are calculated using the set of parameters that maximise the cost function for each value of the cost function parameter.

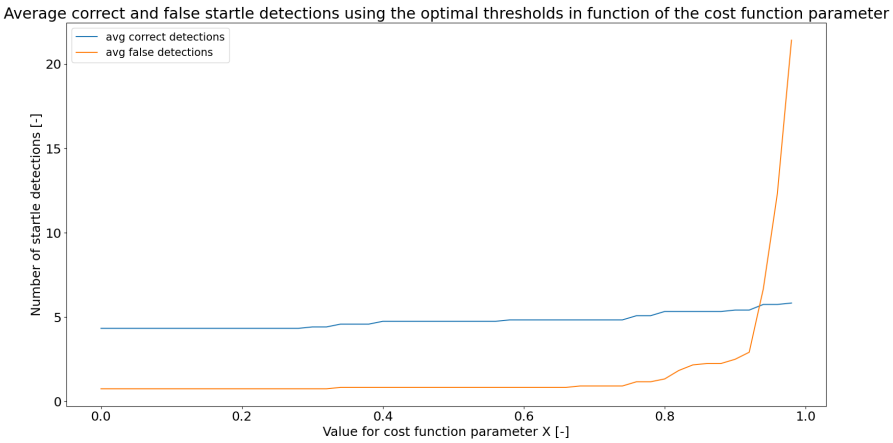
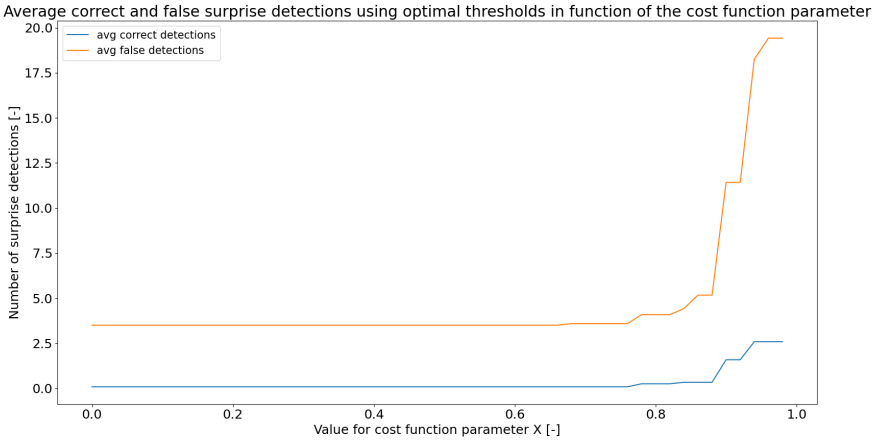
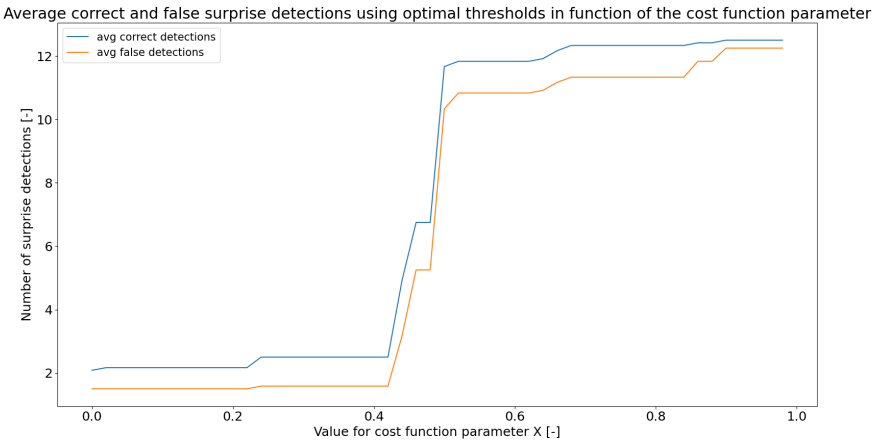


Figure K.1: Startle detections in function of cost function parameter X for verification data set (N = 12)



(a) Only counting rare non-target stimuli as surprising stimuli



(b) Counting rare non-target and frequent target stimuli as surprising stimuli

Figure K.2: Surprise detections in function of cost functino parameter X for verification data set (N = 12)

ICA selection GUI

The GUI used for selecting independent components is depicted in Figure L.1. The first three subplots depict the independent components extracted by ICA. The fourth pane depicts the column averaged of the ICA un-mixing matrix, representing the sign of the back-projection for each component. The lowest pane depicts the averaged EEG signal across electrodes. Bij left-clicking on a component, this component is selected for back-rpojection. IF no components are selected and the plot is closed, no components will be back-projected for this response. After selection of components is performed, a second selection round on

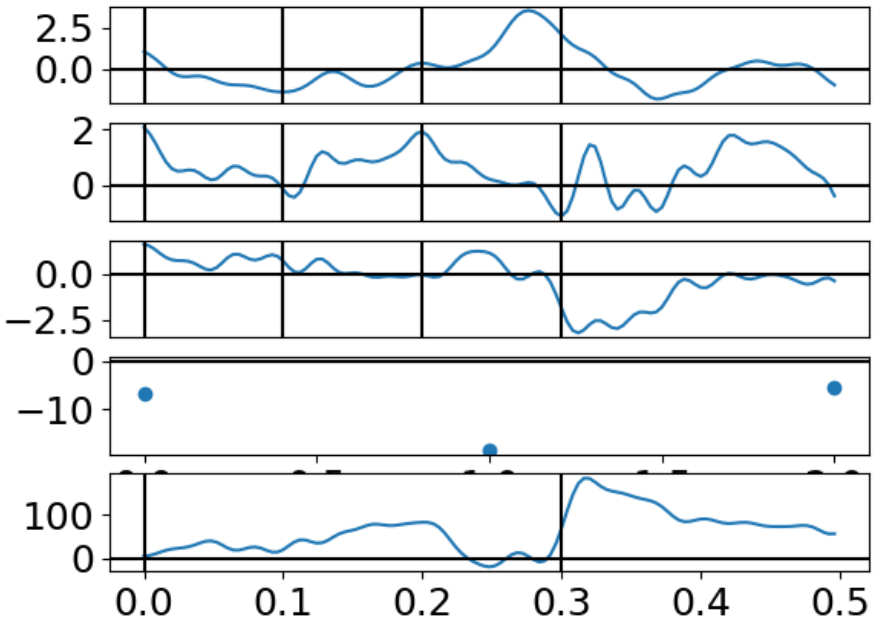


Figure L.1: ICA component selection GUI

the back-projected components needs to be performed. The GUI used for this is depicted in Figure L.2 . By left-clicking on the plot, you select this back-projection as being a P300 ERP, by right-clicking or by closing the plot, you deselect this back-projection as being a P300 ERP.

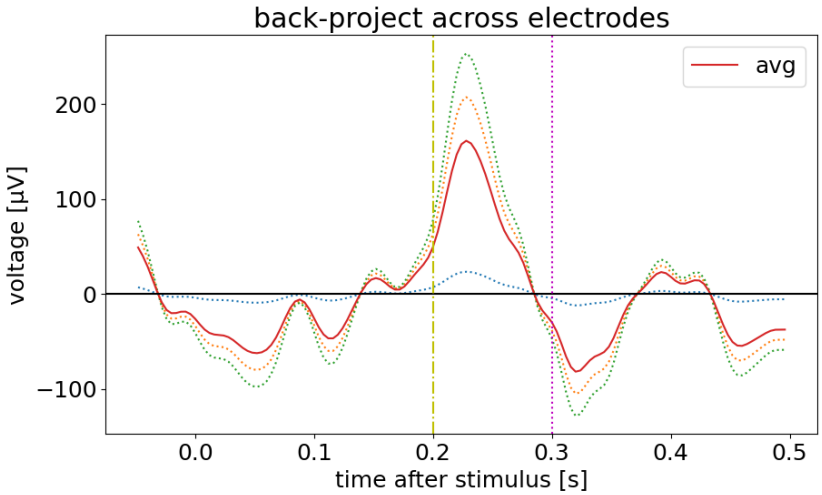


Figure L.2: Back-projection selection GUI

M

Example detection output

When running $startle\_and\_surprise\_detection.py$ with the parameter set as on GitLab, Figures M.1 and M.2 should be outputted. The program should analyse the file named $Data/part\_10\_surp\_seq\_0\_start\_seq2\_trainingFalse.hd$

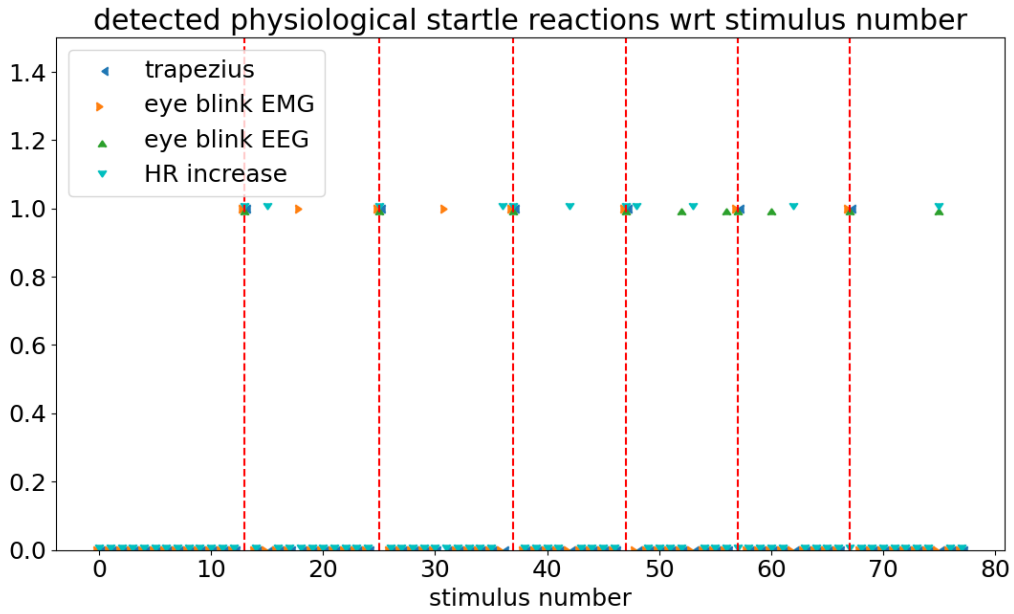


Figure M.1: Example startle detection output

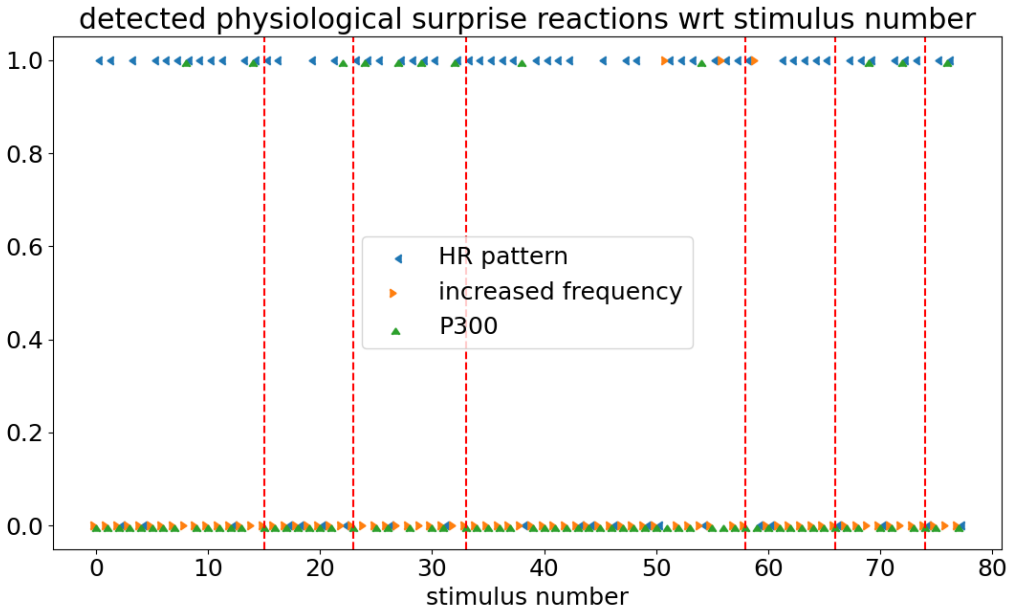


Figure M.2: Example surprise detection output

N

Confusion matrices for detection of separate effects

In the technical report, only the confusion matrices for the validation startle and surprise detections were presented. This appendix provides the confusion matrices for each effect separately. All results are based on the validation set of ten participants.

N.1. Effects related to startle

Tables N.1, N.2, N.3 and N.4 depict the confusion tables for the startle effects.

	Detection	No detection
Startle stimulus	4.7	1.3
No startle stimulus	0.6	71.4

Table N.1: Confusion matrix for EMG trapezius detections

	Detection	No detection
Startle stimulus	4.9	1.1
No startle stimulus	1.7	70.3

Table N.2: Confusion matrix for EMG eye blink detections

	Detection	No detection
Startle stimulus	5.0	1.0
No startle stimulus	0.7	71.3

Table N.3: Confusion matrix for EEG eye blink detections

N.2. Effects related to surprise

Tables N.5, N.6, N.7 and N.8 depict the confusion tables for the startle effects.

	Detection	No detection
Startle stimulus	1.1	4.9
No startle stimulus	4.8	67.2

Table N.4: Confusion matrix for ECG increased HR detections

	Detection	No detection
Startle stimulus	1.9	4.1
No startle stimulus	29.1	42.9

Table N.5: Confusion matrix for EEG delta frequency band detections

	Detection	No detection
Startle stimulus	2.1	3.9
No startle stimulus	23.9	48.1

Table N.6: Confusion matrix for EEG theta frequency band detections

	Detection	No detection
Startle stimulus	5.0	1.0
No startle stimulus	64.4	7.6

Table N.7: Confusion matrix for ECG surprise HR pattern detections

	Detection	No detection
Startle stimulus	1.3	4.7
No startle stimulus	7.9	64.1

Table N.8: Confusion matrix for ECG P300 ERP detections