A Hybrid Brain-Computer Interface using P300 and Steady State Visual Evoked Potentials

by

Matheus Gonçalves Mussi

A thesis submitted in partial fulfilment for the requirements of the degree of

Master of Science

in

REHABILITATION SCIENCE

Faculty of REHABILITATION MEDICINE

University of Alberta

Abstract

BACKGROUND: Children experiencing neurological impairment can experience limitations in their functional abilities. For people with severe physical disabilities, brain-computer interfaces (BCI) are a potential solution to access computers when other assistive technologies prove to be inaccessible. Although BCIs can help individuals accomplish a number of activities, some traditional BCI methods yield insufficient performance to be used in online applications. Hybrid BCI (hBCI) systems aim to improve the system's performance by combining brain signal paradigms, or brain signals with other inputs.

OBJECTIVES: The purpose of this study was to develop and test an EEG-based hBCI system using P300 and steady-state visual evoked potentials (SSVEP) simultaneously, and compare the performance of the developed hBCI against the pure P300 and SSVEP BCI in offline and online scenarios.

METHODS: This study validated the system and potential measures with adults without disabilities. It includes two parts. The system was developed in part 1 with eight neurotypical adults who tested the system at different stages of design. Using the user-centered design, the system was modified based on the volunteers' opinions and the final system was used in part 2. Six different neurotypical adults, divided into two groups, tested the system in part 2. The participants performed six sessions over three weeks, two with each paradigm (P300, SSVEP, hybrid). The second group used a system modified slightly to improve performance. The performed task was programmed so that three targets flickered at different frequencies to generate the SSVEP response and frames appeared semi-randomly to generate the P300 response. The system was evaluated in accordance with Kübler's usability measures of effectiveness, efficiency, and satisfaction. For the effectiveness, three types of accuracies were calculated during the sessions: offline, continuous and selection. For the efficiency, the response time for online sets was measured and the information transfer rate was calculated. For the satisfaction, the NASA TLX questionnaire was used to evaluate the workload of each paradigm.

RESULTS: For group 1, the average selection accuracy for the pure P300 was 83.33%±11.86, for the pure SSVEP it was 49.52% ±17.79, and for the hybrid it was 49.44% ±13.04. For group 2, the average selection accuracy for the pure P300 was 96.39% ±4.29, for the pure SSVEP it was 49.72% ±11.54, and for the hybrid it was 49.44% ±17.31. For group 1, the ITR for the pure P300 was 64.53 bits/min, for the pure SSVEP it was 6.73 bits/min, and for the hybrid it was 6.68 bits/min. For group 2, the ITR for the pure P300 was 111.27 bits/min, for the pure SSVEP it was 6.90 bits/min, and for the hybrid it was 6.68 bits/min. The workload was calculated for each system (0 lowest and 10 highest). The average final workload was 3.27±1.59 for the P300, 5.02±1.22 for the SSVEP and 5.36±1.49 for the hybrid. Post-analysis showed that the lower accuracy on the hybrid was a consequence of the lower accuracy of the SSVEP.

CONCLUSION: The hybrid combination of the P300 and SSVEP did not result in the expected improvement in this study. The attempt to use a short sampling window size of 0.5s might account for the SSVEP's poor performance, which consequently negatively affected the hybrid performance. Although changes made from group 1 to group 2 improved the time response for all paradigms, the changes were unable to sufficiently improve the SSVEP accuracy. Recommendations to increase the accuracy of the system are suggested for future studies.

Preface

This is an original work by Matheus Gonçalves Mussi. The research project of which this thesis is a part, received research ethics approval from the University of Alberta Health research Ehtics Board: “Access to Play through a non-invasive Brain Computer Interface”, Pro00096816_AME2, November 20, 2020.

"If it seem to thee that thou knowest many things and understandest them well enough, know at the same time that there are many more things of which thou art ignorant."

Thomas à Kempis, The Imitation of Christ

Contents

List of Figures

List of Tables

1 Introduction

Children experiencing neurological impairment, e.g. living with cerebral palsy, stroke or spinal-cord injury, can have severe limitations in their functional abilities (Bauer et al., 1979). These conditions can limit the capacity to move, speak and perform other activities independently. Assistive technologies can enable independent participation in activities (Cook & Polgar, 2014). For example, individuals that have lower limb impairment can use power mobility to move to different locations, those with communication impairments can use a variety of communication systems, complete with applications that support computerized voice communication, and those with upper limb impairment can use switches and scanning to access wheelchairs, communication systems, and computers. For people with severe physical disabilities, the use of the brain-computer interfaces (BCI) can be a potential solution when other access methods may prove inadequate. BCI may provide access to power mobility, communication, and play, which are all activities that can improve quality of life (Carelli et al., 2017). BCIs can have different configurations to capture brain signals (Fernández et al., 2014; Luck, 2014; J. R. Wolpaw & Wolpaw, 2012) This thesis focuses on non-invasive technologies for BCI, specifically electroencephalography (EEG) -based systems.

A brain communicates through electrical activity. The outermost layer, called the cortex, is responsible for sensing signals that enter the central nervous system and issue brain signals (Purves et al., 2004). The cortex is traditionally separated into four regions: Frontal, Parietal, Occipital and Temporal. The Frontal region processes high-order executive functions; the Parietal receives and associates somatosensory, visual and auditory inputs; the Occipital mainly consists of the visual processing area; and the Temporal lobes are essential for memory, high-level visual and auditory understanding (Brodmann, 1909). In EEG, BCI input comes from electrodes that capture brain signals on the surface of the scalp. Electrodes are denominated according to the region they are located (F for Frontal, P for Parietal, O for Occipital and T for Temporal) (*Report of the Committee on Methods of Clinical Examination in Electroencephalography*, 1958).

The BCI can be described as a human/technology interface, which interprets brain signals to control other devices, as seen in Figure 1.1. Brain signals are processed in two stages, signal conditioning and classification. The signal conditioning stage aims to transform the EEG raw input, usually through amplification and filtering, so features of the brain signal can be extracted. Most BCI features are temporal, spectral or spatial and more than one can be extracted at a time, creating a feature vector. Features example can be amplitude, latency, and power spectrum density. The feature vector is then used by the classifier to identify patterns across trials and participants (J. Wolpaw & Wolpaw, 2012). The

classifier's goal is to indicate the corresponding target selected based on the features of the signal. In some systems, calibration is required, which is commonly referred to as training. The training uses data collected offline to "teach" the classifier how to interpret new data input into the system. For online applications, a trained classifier is fed with unseen data in real-time. From the classifier, activity outputs are generated, which are used to control the desired activity device.

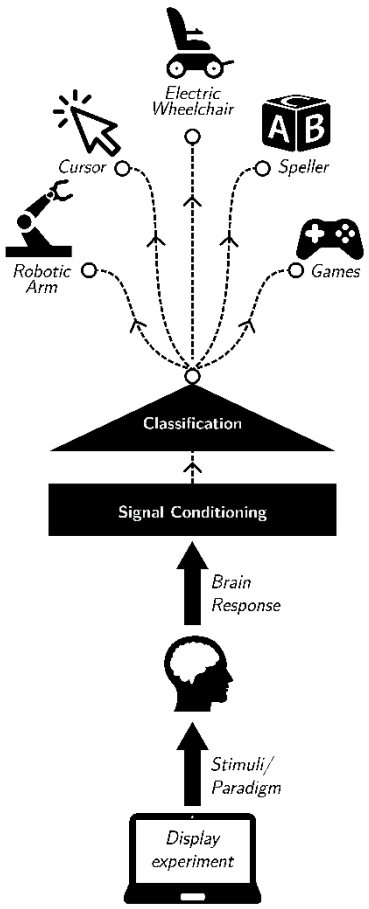

Figure 1.1: Brain-computer interface functioning scheme

Performance is the key factor for patients to decide to assistive systems or not. Huggins et al. (2011) interviewed 61 people with amyotrophic lateral sclerosis that could be potential BCI users. The authors found that the potential users expressed that accuracy is one of the most important parameters and it would only be acceptable to switch from a traditional assistive technology to BCI if the BCI system had classification accuracy of 90% or higher.

Collinger et al. (2013) interviewed 57 veterans with spinal cord injury and found that the most important feature in a BCI system is independent operation. Blain-Moraes et al. (2012) interviewed eight individuals with amyotrophic lateral sclerosis who said using BCI can give them more freedom, but interviewees also

pointed out that “BCI technology in its current form would not be acceptable or appropriate” for daily activities (Blain-Moraes et al., 2012). The main reason for the unacceptability was the fatigue created by BCI. Fatigue can be solved or reduced by improving classification accuracy, which will diminish the number of times the user needs to correct his selection and, consequently, the time needed to perform activities.

Another performance measurement is the Information Transfer Rate (ITR), which determines how much information is transmitted, considering speed and accuracy, and it can be compared across applications (Pierce, 1980).

Although BCIs can help individuals accomplish a number of activities, some traditional BCI methods yield insufficient performance to be used in online applications (i.e. as a real-time system). Researcher have utilized a variety of techniques to attempt to increase BCI accuracy with mixed results due to complex methodological barriers (Batres-Mendoza et al., 2017). Complex algorithmic solutions and deep learning methods have the downfall of requiring powerful processing and in turn driving up time for analysis. Other approaches have looked at ways to improve classification performance using deep neural networks (Borhani et al., 2019; Jiang et al., 2019; Rong et al., 2020), but they needed extensive time offline to process training data and the classification performance was still below the expected standards.

Another limitation of traditional BCI is that they rely on a single input signal (e.g. EEG), single source of stimulus (e.g. auditory, visual, tactile, etc.) or a single brain signal paradigm (patterns), and thus the system has an inflexible human-interface and less information to improve its performance (Zina Li et al., 2019).

Taken together, these limitations are driving hybrid BCI (hBCI) research toward becoming a desirable option. The main goal of hBCI is to improve BCI system performance through multi-modal signal inputs, e.g. combinations of different brain signals, BCI paradigms and/or other external device stimuli (J. Wolpaw & Wolpaw, 2012).

The purpose of this study was to develop and test the software to implement a low-density EEG system hBCI system. Before using BCI with individuals who have disabilities, some authors have first validated the system and potential measures with adults without disabilities (Dovgialo et al., 2018). This strategy was used in this thesis in order to provide a baseline of how the system performs in more controlled conditions. The developed system will be tested with children in future research. It is expected that parameters will need to be adjusted for children, but that the overall system functioning will be similar to the adults. BCI research in general has mainly been tested with adults, and there is a lack of BCI

implementation in children (Kinney-Lang et al., 2016; Mikołajewska & Mikołajewski, 2014). There are challenges regarding children's brain signals that have not been fully addressed. For example, how developmental changes throughout the childhood may affect EEG interpretation and alter acquisition techniques required for pediatric BCI, especially for children with disabilities (Kinney-Lang et al., 2016, 2019). However, because of the potential for BCI to enable children with disabilities to access play and social activities, efforts should be made to develop BCI for them.

After reviewing the research that has been developed around BCI and hBCI geared towards clinical applications, the hBCI study is presented. This study aimed to achieve three main objectives. First, to develop a hybrid-BCI system using P300 and SSVEP simultaneously. Second, to test if the hybrid surpassed single input BCI in classification accuracy and selection time. Third, to gather participants' perceived satisfaction with the system.

It is acknowledged that there are individual preferences about terminology around disability, but person-first language was chosen for this thesis for individuals who have disabilities. Efforts were made to us it in a respectful manner.

2 Literature review

A review of studies using BCI and hybrid BCI (hBCI) was performed to examine the different types of control paradigms used, and the hBCI combinations that have been applied. A brief analysis of the accuracies attained with different paradigms and classification algorithms is presented. The initial search terms used in the IEEE and Scopus databases and in Google Scholar were:

"(Child* *OR* (young adults)) *AND*

(Brain-computer Interface* *OR* BCI *OR* human-machine interface *OR* HMI *OR* human-computer interface *OR* HCI) *AND*

(Electroencephalography *OR* EEG)".

Since few articles resulted from the search, research with adults were also included.

Articles were included for review if they were regarding **clinical applications** (i.e., with an eventual targeted user population of people who have disabilities). Other papers were added based on reference list reviews and suggested related papers.

2.1 Brain-computer interface studies

The articles from the literature review are presented in the following paragraphs according to the paradigm that was used in the studies. A BCI paradigm is the experimental protocol or the set of tasks that elicit a specific type of brain activity (Hwang et al., 2013). In EEG-based BCI's, there are three primary types of brain activity typically investigated: slow cortical potentials, sensorimotor rhythms and evoked potentials.

Slow cortical potentials are invoked (i.e. originated by the person) when individuals up- or down-regulate their own cortical activity. Negative (down-regulating) and positive (up-regulating) slow cortical potentials depolarize and polarize, respectively, the cortical network, causing a brain activity level that can be detected and used for activity outputs. An early protocol example of slow cortical potentials is the contingent negative variation (Walter et al., 1964). This protocol utilizes a "slow-going negative event-related potential" that occurs between a warning and a stimulus that requires a motor response" (Bareš et al., 2007). The main goal is to use this method without feedback; therefore, this modality requires long periods of training to become efficient. To train, the paradigm for new users is associating positive emotions to negative slow cortical potentials and relaxing activities to positive slow cortical potentials (Albrecht et al., 2017).

Sensorimotor rhythms are invoked when the frequency of the signal power changes in the sensorimotor area. These are self-induced, and no external stimuli are needed. Sensorimotor rhythms are movement-related potentials, induced by either executing or imagining movement. Sensorimotor rhythms require a considerable amount of training to be well executed by BCI users (Lotze & Halsband, 2006). The most common paradigm for sensorimotor rhythms is motor imagery (MI), which consists of thinking about or attempting movement without necessarily performing an actual movement. Two different frequency power changes can be seen and classified, the event-related de-synchronization (ERD) and the event-related synchronization (ERS). In ERD in typical developing adults, power decreases in the alpha (8 – 12 Hz) and beta (18 – 26 Hz) bands before the MI is performed (frequencies may change at different developmental stages) (Lazarou et al., 2018); in ERS, power increases in the beta band after the end of the MI. These power changes have been used to control cursor movements, game applications and external devices (Lazarou et al., 2018).

Finally, evoked potentials are natural brain responses to specific external stimuli. Two paradigms typically associated with this BCI paradigm are the P300 and Steady-State Evoked Potentials (SSEP). The P300 consists of a peak in the brain signal with a latency of 300 ms after the stimulus. P300 can be evoked through visual, auditory or tactile stimuli. In the P300 paradigm, stimuli are presented in a random order, with the participant needing to attend to the desired target stimulus. Each time the attended stimuli is activated, a P300 evoked potential is generated. For P300, the classification process aims to identify the characteristic signal peaks evoked by P300 stimuli. Usually, less training than MI is required for new users (Hwang et al., 2013) and the pattern is consistent across individuals. Also, independently of how many targets are presented, the selection is less complex because targets can be targets or non-targets only (multiple choices are reduced to a binary selection). But this method usually has insufficient ITR and has lower classification accuracy compared to other evoked potentials, e.g. SSEP (Lazarou et al., 2018).

SSEP paradigms can be visual, Steady State Visual Evoked Potential (SSVEP), or auditory, Steady State Auditory Evoked Potential (SSAEP). SSVEP are elicited when flickering visual stimuli are presented at consistent (e.g. ‘steady’) frequencies. SSAEP works similarly to SSVEP, but differing sound frequencies are used (i.e. playing a sine wave sound at the given frequency). When the individual focuses on a particular stimulus, pyramidal cells resonate at the same frequency, and through power frequency analysis it is possible to distinguish the desired selection (Lazarou et al., 2018). Nevertheless, it is important to consider what frequencies are used. Some frequencies for visual stimulus between 12-25 Hz may induce seizure in people with photosensitivity (Fisher et al., 2005; Okudan & Özkara, 2018).

Another example of evoked potentials is the so-called error potentials (ErrPs) category. They are elicited by the brain when an error happens, i.e. when the output yielded by the system is not the same as that desired by the user. Some systems use the error potentials to correct misclassification and increase the accuracy, as in Yousefi et al. (2019) where the accuracy increased from 60% to 67% in real-time trials using ErrPs.

2.2 Hybrid brain-computer interface studies

Hybrid BCI can be achieved through three primary combinations of resources: data which joins multiple **brain patterns**, data which combines **multisensory stimuli** and data composed of **multiple signals sources**. Utilizing multiple brain patterns together is in effect combining two different types of brain activities, e.g. SSVEP and P300. In Pan et al. (2014), they combined SSVEP and P300 paradigms to detect awareness in several patients with brain injuries. A picture of a known relative appeared on the screen alongside with a picture of an unfamiliar face. Both pictures flickered at different frequencies (SSVEP component) and a white frame randomly appeared around the pictures, one at a time (P300 component). While the patient focused on the familiar face, he was also instructed to count how many times the frame showed around the picture. The average accuracy was 72.01%, with some participants reaching as high as 100% and 96.67% during trials. Yin et al. (2015a) also combined SSVEP and P300 paradigms for use in a speller. The average results from online trials report an accuracy increase from 91.33% to 95.18%, with an increase of ITR to 50.14 bits/minute, compared to the 47.14 bits/minute previously reported.

Other brain activity combinations, as in in Zuo et al. (2019), combined MI and P300 in a task where participants had to choose between two Chinese symbols on a screen. In their study, they classified both the P300 and MI but because P300 classifiers have high accuracy independently, compared to the steeper learning curve of MI, the overall output was dominated by the P300 classification only. However, the authors used the initial P300 outputs to help inform the MI classification. Once the MI classification output surpassed what the authors determined as a reliability threshold, the P300 and the MI classification were compared. If both outputs coincided, the selected output was maintained; if they diverged, and the reliability threshold was surpassed, the MI output was selected; otherwise, the P300 output was chosen. The overall average online accuracy of the system was 93.94 ± 5.19% using P300 + MI, which was higher compared to P300 (91.25 ± 9.04%) and MI (81.61 ± 8.79%) methods alone.

Multisensory stimuli combinations evoke reinforced brain signal patterns through different sensory modalities (for example, audio-visual or visual-tactile). Carmona et al. (2020) combined visual and auditory stimulation through SSVEP and SSAEP paradigms. In the experiment, frequencies of 37, 38, 39

and 40 Hz were used to stimulate the participants visually and aurally (flashing and beeping frequency). Trials consisted of visual only, auditory only and joint visual-auditory modalities. The SNR increased considerably when using the visual-auditory modality (from 1.1 to 1.4, on average). The highest accuracy reached by one of the participants was over 95%, the average accuracy of the classified data from the electrode Oz was between 70 and 80% and the results ranged from approximately 48% to 95%. Moreover, they found that the results of classification from electrodes on non-hair positions (i.e., Tp9 and Tp10) were statistically similar to the results of classification from the occipital electrode (Oz). Yin et. al (2015b) combined tactile- and auditory-P300 modalities to create a multisensory hBCI. Four pairs of motors were attached to the participant's waist and four computer speakers were laid out in a circle around the participant. Each speaker issued a voice saying its number as an auditory stimulus, and the corresponding motor vibrated accordingly. The average accuracy using this hBCI was of 88.67%, and the ITR of 10.77 bits/minute. Thurlings et al. (2014) also did a multisensory experiment in which they used P300 with visual and tactile stimuli. An actuation pair, composed of a small vibrating motor and an LED, was attached to a finger on each hand of the participant. The classification accuracy using only visual, only tactile and visual-tactile stimuli modalities was compared. The visual-tactile modality gave the highest accuracy rates in online trials, reaching an average of approximately 85%, surpassing the visual only or tactile only modalities, which had an approximate accuracy of 70%, as estimated from the published graphical data.

Finally, an hBCI can combine multiple signal sources, such as the BCI and another type of interface like eye gaze or switch input. This combination of signal sources aims to extract signal paradigms from the same event simultaneously or from sequential events that are combined to accomplish an activity. Saravana and Reddy M. (2018) extracted signals simultaneously from an SSVEP paradigm and video-oculography (VOG) (i.e. eye-tracking based on computational vision approach) system. In the experiment, participants selected letters on a virtual keyboard. While the letters flickered according to the stipulated SSVEP frequencies, the VOG system tracked participant's gaze, allowing the system to increase its accuracy. Both inputs were combined and gave an online average accuracy of 94.99% with ITR of 82.78 bits/minute. In Huang et al (2019), multiple signals were combined sequentially using an hBCI based on MI and electrooculography (EOG), a technique that measures electrical signals related to eye movement. The participants selected among nine different options on a small screen to move a wheelchair. All options were cyclically highlighted one at a time, waiting for the user to blink once to indicate a selection. When the user blinked, the highlighted option would start flashing in the panel for confirmation. To confirm the selection, the user needed to raise their eye-brows. Then the participant used MI to move the wheelchair in the chosen direction. In online sessions, every 0.2 s of MI input were

compared to the offline training data. Depending on the similarity of the signal with one of the three states (right MI, left MI or idle), the 0.2 s input received a score. If the score surpassed the stipulated right- or left-threshold, the wheelchair would start moving towards the selected direction. The system could also control a robot arm, mounted on the wheelchair, which moved along pre-calculated trajectories. The user used EOG to select between two different bottles and the robot automatically brought the target bottle to the participant's mouth. MI average online accuracy for the wheelchair was 88% and EOG was 96.2%.

There are other hBCI that are not EEG-based. Schudlo and Chau (2018) combined three different signals from near-infrared spectroscopy (NIRS) BCI. NIRS can measure the concentration of oxygenated, deoxygenated blood and total hemoglobin in the brain. They used those three signals to compose their system. Also, Faress and Chau (2013) combined functional NIRS and functional transcranial Doppler ultrasonography to try improving the accuracy of the system. However, the techniques used in these projects are not useful for EEG-based BCI systems, thus, they will not be discussed further.

2.3 Analysis and Comparisons of BCI and hBCI Studies

All of the papers presented in the following analysis are listed in Appendix 2.

2.3.1 Highest BCI accuracies

The papers in Appendix 2 were examined to understand the impact of the paradigms and the combination of paradigms on accuracy. The eight highest accuracies are covered in detail and listed in Table 2.1. Zhang et al. (2017) used P300 to allow participants to select among cups they could drink from. The selection activated the robotic arm. The screen displayed four options to the participant: three cups that could be selected and an option to put the drink back on the table. After the selection was classified, the robotic arm picked up the selected cup and brought it to the participant's mouth. This system's P300 classifier gave an accuracy of 97.5%. The study of Choi et al. (2018) proposed a method for users to play chess through the combination of SSVEP and EOG, where users could select the final-position for the chess piece through EOG and confirm the selection with SSVEP. They achieved accuracy results of 85.8% for the SSVEP step and 96.3% for the EOG step. The authors also examined a joint P300 and EOG paradigm, but report this combination had a lower accuracy (<70% for the P300). They hypothesize the EOG and P300 combination was not as accurate because of signal interference, as signals captured by EOG electrodes are also captured by P300 electrodes and vice-versa.

2.3.2 Paradigm and Activity

The review showed that certain paradigms were used for certain activities. The relationship between paradigm and activities and their influence on successful BCI is further examined in this subsection.

Table 2.1 - Highest accuracies from the studies in Appendix 2

Study	Classification Accuracy [%]	N. of Classes	Paradigm
(Carmona et al., 2020)	~95	2	SSVEP + SSAEP*
(Z. Zhang et al., 2017)	97.5	2	P300
(Yin et al., 2015a)	95.18	2	P300 + SSVEP
(Saravanakumar & Reddy M., 2018)	94.99	2	SSVEP + VOG
(Zuo et al., 2019)	93.94	2	MI + P300
(Huang et al., 2019)	88	3	MI + EOG
(Yin et al., 2015b)	88.67	2	Tactile P300 + Auditory P300*
(Choi et al., 2018)	85.8	2	SSVEP + EOG
(Thurlings et al., 2014)	~85	2	Tactile P300 + Auditory P300*

*Note. *:Paradigms with multisensory stimuli.*

2.3.2.1 Motor Imagery

In general, sensorimotor paradigms (MI) are used in process-control activities. Table 2.2 outlines the average accuracies for each study in this section. Reported MI applications include moving a cursor from left to right on a monitor (Cincotti et al., 2008; J. Z. Zhang et al., 2019), moving robotic arms (Kim et al., 2019; Meng et al., 2016), and wheelchairs (Huang et al., 2019). Other applications have used MI as an target selector (Zuo et al., 2019), where imagining a left-movement selected the left option and right-movement selected the right option on a screen. Meng et al. (2016) describe a MI BCI system which moves a robot in 3-dimensional coordinates utilizing a sequential combination of "low dimensional controls" (i.e., using only two commands at a time). The first control moved the robot arm in a 2-dimensional plane parallel to the table, where right/left MI moved the robot right and left respectively. For the second control, they categorized MI activity of simultaneous hand movement or resting state to allow the robot to move backwards (down) or forward (up), respectively depending on task requirements. Zhang's et al. (2019) results were expressed through Cohen's Kappa average.

Table 2.2 - MI studies from the studies listed in the Appendix 2

Study	Activity	N. of Classes	Result
(Cincotti et al., 2008)	Cursor Movement	2	Acc: 86.75%
(Meng et al., 2016)	Robot arm in 3D space	2, 4	Acc_2: ~95%, Acc_4: ~85%
(Cho et al., 2017)	Feedback	2	Acc: 67.46%
(J. Z. Zhang et al., 2019)	Car and Cursor Movement	2	Kappa: 0.46*
(Kim et al., 2019)	Robot arm in 2D space	2	Acc: 57.37%
(Huang et al., 2019)	Wheelchair	2	Acc: 88.00%

Note. Meng et al. (2016) had two phases on their study, using 2 and 4 options for participants to select.:Cohen's Kappa cannot be translated to accuracy [%] because they are not directly comparable, Acc: Accuracy.*

2.3.2.2 Evoked Potentials

Evoked potentials were used in goal-selection activities. Using the SSVEP paradigm, participants selected among options in game menus (Choi et al., 2018), commands to move a speller cursor (Ehlers et al., 2012) or characters on a keyboard (Saravanakumar & Reddy M., 2018; Yin et al., 2015a). P300-based BCI were used to control a semi-automated robot (Z. Zhang et al., 2017), to play games (Choi et al., 2018) or to make selections to test the system's accuracy (Yin et al., 2015b; Zuo et al., 2019).

Table 2.3 summarizes the studies' average accuracies for evoked potentials paradigm. Ehlers et al. (2012) presented accuracies for different age ranges. It is important to note that the main task of the experiment was a spelling task using a 4-arrow cursor that might have influenced children to do worse than adults. From the tested frequencies in this study, the authors presented the medium range frequencies (13-17 Hz) for the SSVEP that gave the highest results for all groups. In group 1, the average age was 6.73 years old (y.o.) and the accuracy was 58%, in group 2 the average age was 8.08 y.o. and the accuracy was 53%; in group 3 the average age was 9.86 y.o. and the accuracy was 75% and in group 4 the average age was 22.36 y.o. and the accuracy was 78%.

Table 2.3 - SSEP studies from the studies listed in the Appendix 2

Study	Paradigm	Activity	Accuracy [%]
(Ehlers et al., 2012)	SSVEP	Command Selection	58; 53; 75; 78*
(Choi et al., 2018)	SSVEP	Chess Game	85.8
(Choi et al., 2018)	P300	Chess Fame	67.8
(Saravanakumar & Reddy M., 2018)	SSVEP	Virtual Keyboard	94.99
(Carmona et al., 2020)	SSVEP + SSAEP	Feedback	~75
(Yin et al., 2015a)	SSVEP + P300	Virtual Keyboard	95.18
(Yin et al., 2015b)	Tactile P300 + Auditory P300	Feedback	88.67
(Thurlings et al., 2014)	P300*	Feedback	~85
(Z. Zhang et al., 2017)	P300	Command Selection	97.5

*Note. *Accuracy for Group 1 (6.73 y.o), Group 2 (8.08 y.o), Group 3 (9.86 y.o.) and Group 4 (22.36 y.o), respectively.*

2.3.3 Classifiers

In the examined studies, there was a **relationship between paradigms used and classifiers used**. There are some features that provide more information depending on the type of brain activity. For example, common features extracted from SSEP paradigms have rich spectral information, and spatial information depending if it is visual, tactile or auditory, but poor temporal information. Linear discriminant classifiers are better suited for binary outputs, like P300 outputs (target versus non-target), because they are based on statistical regression and map the data into new spaces to maximize data separability (J. Wolpaw & Wolpaw, 2012). CCA classifiers use sinusoidal waves as reference signals for classification, making it a powerful tool to identify spectral components, present in SSEP modalities (Nakanishi et al., 2015). Power frequency classifiers look for changes in specific frequency ranges, and MI has defined bands of frequency that change (alpha and beta, especially) (J. Wolpaw & Wolpaw, 2012). However, although some paradigm-classifier combinations are more common, in an effort to surpass previous reported accuracies, many authors try different classifier combinations and new methodologies.

The Paradigm-Classifier relationships are shown in Table 2.4. Most studies reported here used classifiers that fall under the linear discriminant category of classifiers, which are variations of the Fisher Linear Discriminant Analysis (LDA), such as Bayesian LDA (BLDA), Stepwise LDA (SWLDA) and regularized LDA (rLDA). From the analyzed studies, a relationship classifier-paradigm can be seen for both hybrid and single-paradigm BCI, such as LDA-P300 and CCA-SSVEP, as mentioned by Wolpaw & Wolpaw (2012) and Nakanishi (2015). Cho et al. (2017) was an exception, using LDA to classify MI, as well as

Zuo et al. (2019), using BLDA to classify both MI and P300. Cincotti et al. (2008) used statistical analysis (to assign significance to the frequencies for different conditions) to classify MI.

Multisensory stimuli hybrids used the same type of classifier for both stimuli. Thurlings et al. (2014) and Yin et al. (2015b) combined visual and tactile P300 stimuli, and had for both paradigms a linear discriminant classifier. Carmona et al. (2020) applied a CCA classifer for both SSVEP and SSAEP stimuli.

Table 2.4 - Classifier comparison among studies listed in the Appendix 2

Type of Classifier	Study	Classifier	Paradigm	Type of System	Hybrid Nature	Accuracy [%]
LDA	(Z. Zhang et al., 2017)	BLDA	P300	BCI	N/A	97.5
	(Yin et al., 2015a)	SWLDA	P300	hBCI	P300 + SSVEP	95.18
	(Zuo et al., 2019)	BLDA	MI + P300	hBCI	MI + P300	93.94
	(Yin et al., 2015b)	BLDA	P300*	hBCI	Tactile P300 + Auditory P300	88.67
	(Thurlings et al., 2014)	SWLDA	P300*	hBCI	Tactile P300 + Visual P300	~85
	(Choi et al., 2018)	LDA	P300	hBCI	P300 + EOG	67.8
	(Cho et al., 2017)	LDA	MI	BCI	N/A	67.46
	(Yousefi et al., 2019)	rLDA	Evoked Pot.	BCI	N/A	67
CCA	(Yin et al., 2015a)	CCA	SSVEP	hBCI	P300 + SSVEP	95.18
	(Choi et al., 2018)	CCA	SSVEP	hBCI	SSVEP + EOG	85.8
	(Carmona et al., 2020)	CCA	SSEP*	hBCI	SSVEP + SSAEP	~85
r^2	(Cincotti et al., 2008)	Statistical Analysis	MI	BCI	N/A	86.75

*Note. Paradigms with * mean they were multisensory stimuli. Acronyms: LDA – Linear Discriminant Analysis, BLDA– Bayesian LDA, SWLDA – Stepwise LDA, rLDA – Regularized LDA, CCA – Canonical Correlation Analysis*

2.4 Summary of hBCI literature

In this review, studies related to BCI and hBCI systems were reported and discussed. This section discusses further the overall findings reported.

The hypothesis that hBCI in general have a better accuracy than single-paradigm BCI is supported, as the reviewed systems with hybrid combinations tended to have a higher accuracy. Five of the eight selected papers with **hybrid** systems reported surpassing the **90%** minimum classification accuracy standard indicated by Huggins et al. (2011). Out of the highest accuracy studies, four studies used the SSVEP

paradigm as a component in the hybrid system, four used P300 and two used MI. Except for Carmona et al. (2020), all studies ran online trials.

Although many studies state that their final goal is to implement this technology for people with complex needs, **few of them actually tested their systems with people with disabilities**. From the highest accuracies studies in Table 2.1, participants were between 21 and 37 years old; the average number of participants was approximately 10.67±4.36. In the studies where the same system was used by individuals with and without disabilities, accuracy results for individuals with disabilities were always lower. The system of Cincotti et al. (2008) had an accuracy of 86.75% for people without disabilities, but 66.57% for people with disabilities. The system of Pan et al. (2014) had an average accuracy of 95% for individuals without disabilities, while for people with complex physical needs it was 60.52%. Kim et al. (2019) tested their system with people with disabilities, giving an accuracy of 57.37%.

Most reviewed studies chose their classifier and paradigm based on the final activity. In general, there are some paradigms that are more suitable for certain activities than others and some classifiers that are more suitable for certain paradigms, although this is not compulsory. Sensorimotor paradigms were mostly used for process-control activities such as movements, displacement and adjustment of intensity of activity. There are more robot-related activities using sensorimotor paradigms than evoked potentials. Similarly, evoked potentials paradigms were mostly used for activities that require the selection of targets, mimicking a button push.

As noted in Hwang et al. (2013), the number of **MI-based BCI studies has decreased** and the majority of studies found in this review were mainly using P300 and SSVEP paradigms (as seen in the table of Appendix 2). This trend is potentially due to the lesser amount of time needed for training when using evoked potential paradigms, as compared to typical MI paradigms.

2.5 Discussion of implementation for children

The lack of systems developed and tested for and with children (Kinney-Lang et al., 2016; Mikołajewska & Mikołajewski, 2014) was reflected in this review: only 3 out of the 17 studies listed in the Appendix 2 had children as participants. Some studies that included children reported a decay in performance in comparison to adults, while others did not show a substantial change. Ehlers et al. (2012) indicated an inverse relation between age and accuracy. But the spelling task in the study could have been a factor that contributed to the performance reduction for younger participants. Cincotti et al. (2008) and Pan et al. (2014) included younger participants in their studies (16 y.o.), but they did not have significantly lower accuracies than older participants (25-44 y.o. for Cincotti et al. and 19-70 y.o for Pan et al.). More studies

would be beneficial for the field, since the difference in BCI performance for adults and children is still inconclusive.

The challenges that will arise when implementing BCI for **children will need to be considered**. One challenge commonly explored is how EEG signal features can change with age and development (Kinney-Lang, 2018). Matsuura et al. (1993) analyzed the dominant resting EEG frequency (idle) and noticed that it shifts with age. In children around 7 years it is 8 Hz and in adults it settles at 10 or 11 Hz (Ehlers et al., 2012). The authors also found that the lower the frequency of SSVEP stimulation, the harder it was to identify that frequency when dealing with younger participants. It will be important to consider the population age in future studies so that paradigm setup and pre-conditioning can be adjusted to ensure the best accuracy possible is being reached (Kinney-Lang et al., 2016).

The trend that was observed in the use of paradigms, where most researchers opt to use evoked potentials over sensorimotor rhythms, could be beneficial for implementing BCI for children, since evoked potentials are easier to adjust for each individual's signal features. Also, since evoked potentials require less training, as pointed by Hwang et al. (2013), children may be less frustrated than with other paradigms that require numerous calibration sessions.

3 Methodological approach

The thesis has three objectives: 1) to develop a hybrid-BCI system with P300 and SSVEP; 2) to test if the hybrid surpasses single input BCI in classification accuracy and selection time; and 3) to gather participants' perceived satisfaction with the system.

The chosen BCI paradigms for the proposed hybrid system were the P300 and the SSVEP. This choice was based on the fact that both paradigms are endogenous behaviours, meaning they are naturally elicited by the brain when specific stimuli are presented (Hwang et al., 2013). This leads to less training needed for the system to classify most of the trials correctly, compared to exogenous behaviours which require learning to elicit specific brain signals (Abiri et al., 2019; Donchin et al., 2000). One study that attempted a similar hybrid combination was Pan et al. (2014). They used 10s of signal for processing to make sure the patients could recognize the stimuli. The experiment was online and patients achieved accuracies from 64 to 78%.

We wanted to make the hBCI not only accurate but also quick in determining the final output. Therefore, a 0.5 second sample window from stimulus onset was adopted for both the P300 and the SSVEP paradigm. Some studies use shorter sampling windows for P300 (Cecotti et al., 2010) but SSVEP studies typically use from 1.5 – 2s (Carmona et al., 2020; Ehlers et al., 2012; Pan et al., 2014). The use of a 0.5 s window was expected to be possible since the used frequencies had enough frequency gap to be readable, as determined by the minimum frequency resolution outlined in Equation (3.1) (Harris, 1978; Reyes & Forgach, 2016):

$$\Delta f = \frac{f_s}{N} \qquad (3.1)$$

where f_s is the acquisition frequency (250 Hz) and N is the number of samples (125 samples, given 0.5s). That creates a frequency resolution of 2 Hz. The used frequencies were 6, 10 and 15 Hz, giving a minimum frequency gap of 4 Hz. Additionally, the highest stimulus frequency respected the Nyquist theorem (Diniz et al., 2002).

3.1 Research Design

A user-centred design (UCD) approach was used in this study, as suggested by Kübler et al. (2014) for BCI evaluation. The approach considers the usability of the system determined as effectiveness, efficiency, and user satisfaction. For the effectiveness component, accuracy was used. For the efficiency component, target selection time and information transfer rate (ITR) were used. For the satisfaction component, volunteers' opinions were collected and their ratings on a workload measurement tool were gathered.

The system was tested in two parts. In part 1, the system and interface were developed with in-lab trials, the functionality of the system was validated, and potential enhancements were considered while gathering information about volunteers' satisfaction. Volunteer opinions from part 1 helped to improve the interface and system performance, while informing on and providing practice for the experimental protocols applied in part 2. Part 2 included trials with BCI naive participants. The effectiveness of the system was measured through the accuracy, the ratio of correct classifications among all trials. Accuracy, is defined by Equation (3.2),

$$ACC\% = \frac{TP + TN}{All\ Samples} \times 100 \quad \textbf{(3.2)}$$

in which *TP* is the number of true positives, which is when the positive outcome is predicted correctly by the classifier, *TN* is the number of true negatives, which is when the negative outcome is predicted correctly by the classifier and *All Samples* encompasses all the classifications made. The efficiency of each paradigm was measured by the time each online trial took to make a selection and the ITR. Since each trial took 0.5s, the selection time was calculated based on how many trials were needed for the system to decide on the final classification. ITR is the metric that determines how much information is transmitted, considering speed and accuracy, and it can be compared across applications (Pierce, 1980). It is given in bits/minute. The ITR was calculated with the Equation (3.3) found in Wolpaw and Wolpaw (2012):

$$ITR = m \times \left(\log_2 N + P \log_2 P + (1 - P) \log_2 \frac{1 - P}{N - 1}\right) \quad \textbf{(3.3)}$$

where m is the number of trials per minute, N is the number of commands and P is the accuracy of the system. Due to the logarithmic nature of the ITR calculation, the ITR increases exponentially when the accuracy gets closer to 100%. The satisfaction with the system was evaluated through the NASA TLX questionnaire (Hart & Staveland, 1988), a tool used to assess workload. In other words, satisfaction was evaluated as the participant's agreeability with the perceived effort required to accomplish the task. More details are provided in the *Data collection and analysis* section.

3.1.1 Part 1

Eight volunteers (2 men and 6 women) tested the system at different points of its development. They all tested the acquisition system by itself, gave opinions about the interface, tried each of the pure paradigms and the hybrid paradigm and had data collected for offline processing. Only two tested the system's real-time selection feedback mechanism (i.e. visual feedback about the system's chosen target based on the

participant's input) but no online data was saved. The number of sessions per volunteer varied from 1 to 6.

3.1.1.1 Materials

3.1.1.1.1 EEG Hardware

The EEG acquisition was done using OpenBCI Cyton, which had eight 24 Bit channels and an acquisition rate of 250 Hz. The OpenBCI cap contained silver-plated wet electrodes with electro-gel.

3.1.1.1.2 Initial Interface

The initially designed interface had three flashing squares, for the SSVEP stimuli, and an outline frame that appeared outside of the different squares in pseudo-random order. The frequencies used were 6, 10 and 30 Hz. Based on Floriano et al. (2018), the colours used to create the flashing effect were green and red, to give maximum amplitude in mid-range frequencies from 15-25 Hz. All squares used the same colors for simplicity. No sounds were issued in-between trials or at the end of the experiment. Each square had, approximately, one quarter of the height of the screen. Figure 3.1 is a representation of the interface design.

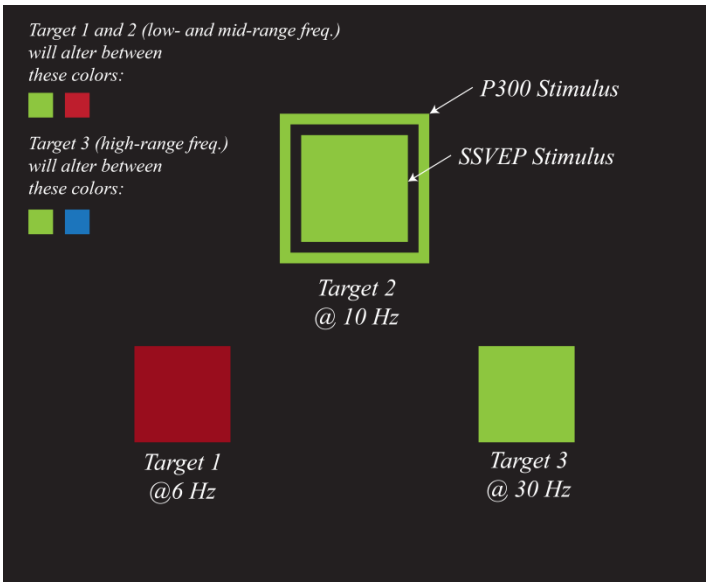

Figure 3.1: Initial interface designed to stimulate SSVEP and P300 responses in users.

3.1.1.2 Procedure

Each session had several training sets and, eventually, one online set. The experiment routine started with a three second count-down to prepare the participant. Then to cue the desired target, its colour changed to orange for three seconds. After the cue, the targets went back to their original colours and the stimuli was presented (depending on whether it was P300, SSVEP or hybrid). In the training sets, the squares flickered uninterruptedly for 10.5 seconds (3 squares × 7 repetitions per square = 21, 21×0.5s = 10.5s)

while the P300 frame appeared semi-randomly around squares for 0.5s with a duty cycle of 80%, i.e. 0.4s on, 0.1s off, in online sessions, the stimulation period was variable because stimuli were presented only until the selection function returned a final answer.

In the training set, after the stimuli, a new cue was presented. The training set had no feedback about what was selected and served exclusively to train the classifiers. This set had a constant number of runs, since it was for the classifier generation. In the online set, feedback about what the selection function chosen was presented for three seconds after the trial ended before a new cue was presented.

For the pure P300 paradigm, the volunteer was instructed to fixate their gaze on the cued square and count how many times it was outlined until a new cue was presented. For the pure SSVEP paradigm, volunteers were instructed to fixate their gaze on the middle of the cued square until a new cue was presented. For the hBCI, the volunteer was instructed to fixate their gaze on the middle of the square and count how many times it is outlined until a new cue was presented.

For the classification of both paradigms, Linear Discriminant Analysis (LDA) classifiers by Sci-Kit Learn library were used (Pedregosa et al., 2011). The solver was set to singular value decomposition and no shrinkage.

The interface and experimental routine were programmed from scratch. The Psychopy library was used for the creation of the visual interface (Peirce et al., 2019). BrainFlow was used to interface the headset and the script (BrainFlow, ©Andrey Parfenov). The code is all based in Python 3 with technical details provided in a previously published extended abstract (Mussi, 2021). (See Appendix 1)

3.1.1.3 Results

The signal conditioning methodology used in part 1 was the same as described below, in part 2. Some of the later results of part 1 are shown in the Table 3.1 below. The results were processed offline using the same algorithm and electrodes placement that were subsequently used in part 2 below. The average accuracies were generated via the 10-fold cross-validation. The stimuli presented were hybrid. Not all participants went through the same number of sets due to the informal nature of the sessions.

Table 3.1 - Data from last sessions in Part 1

Volunteer	Set	P300	SSVEP
V1	1	72.0±12.8%	41.3±11.7%
	2	74.6±8.7%	32.3±7.8%
	3	70.4±7.0%	24.4±6.7%
	4	74.5±9.9%	33.4±10.1%
	5	76.8±12.5%	40.7±10.2%
V2	1	78.3±6.4%	28.8±7.1%
	2	83.1±9.6%	34.9±12.9%
	3	75.2±11.5%	40.1±9.6%
	4	83.1±8.4%	53.0±5.1%
V3	1	82.0±6.4%	78.3±8.1%
	2	75.2±9.4%	71.5±10.8%
	3	81.6±11.6%	68.7±4.7%
	4	85.0±6.9%	55.0±8.0%

After doing the trials with the volunteers, some modifications were made to the interface based on their opinions, and to improve system performance.

- Initially, the experiment schedule proposed two sessions per day, with one session of each of the pure paradigms or two of the hybrid, but volunteers complained about it being too much time focusing on the screen. Most volunteers were exhausted by the end of the first session, so the schedule was adapted to have only one session per day.
- The initial frequencies used had a problem with harmonics (6, 10 and 30 Hz). The 10 Hz frequency created a second degree harmonic at 30 Hz, which induced the classifier to emit more false positives for 30 Hz when the real frequency was 10 Hz. Early on in the testing, the set of frequencies was changed to 6, 10 and 15 Hz, as used previously in the literature (Saravanakumar & Reddy, 2018).
- The usage of green and red flashing to attain maximum signal amplitude was no longer needed, since the newly adopted frequencies fell below the 15-25 Hz mid-range for green and red. Also, the users did not like the flickering between green and red since they said it "blurred their view" more easily. The squares were changed to alternate between black and white for maximum contrast, as commonly used in the literature (Hsu et al., 2016; Yin et al., 2015a).
- Another modification was the addition of a black cross in the centre of each square to help volunteers focus better at the flashing squares.

- By adding the cross, the SSVEP classification improved but the P300 worsened, probably because the cross narrowed the peripheral vision, making the P300 less visible. To accommodate for that issue, the squares were scaled down to about an eighth of the screen height.
- Sounds were added to the end of each trial to help volunteers realize when a new trial was starting.

3.1.2 Part 2

This part was done with two groups of participants. Each group had three participants, and some minor adjustments were made in an attempt to improve the results from group 1 to group 2. Additionally, due to noisy conditions the experiment location was changed for group 2. The specific adjustments will be described in their respective sections.

3.1.2.1 Participants

Six adult men without disabilities and no experience with BCI were recruited from the University of Alberta (Table 3.2). Participants were between 18 and 29 years old (avg. 22.167±3.764), had normal or corrected to normal vision and were right-handed. The study acquired ethics approval from the Research Ethics Office (Pro00096816) and participants signed an informed consent form before proceeding with the study. No participants had any conditions that made them prone to seizures or photosensitivity. Participant 2 had a concussion five years before the experiment. Participants 1 to 3 were part of the first group of participants. Participants 4 to 6 were part of the second group of participants.

3.1.2.2 Setting

The location of the sessions was the assistive technology (AT) lab at Corbett Hall, University of Alberta, for group 1, and at the Eye Gaze lab at Corbett Hall, University of Alberta, for group 2.

Table 3.2 - Participants recruited for part 2

Group	Participant	Age	Additional Information	Experiment Location
1	P1	21	-	AT-Lab
	P2	23	Concussion (18yo) Prescription glasses (-2.25, -1.5)	
	P3	20	Prescription glasses (0.5)	
2	P4	29	-	Eye Gaze Lab
	P5	18	Prescription glasses (1.6)	
	P6	22	Contacts (-3.5, -4)	

3.1.2.3 Materials

3.1.2.3.1 Interface

On a computer display, three squares with white centre areas flashed at different frequencies for the SSVEP component, and an outline frame appeared around the squares one at a time in a pseudo-random order for the P300 component (see Figure 3.2). For the SSVEP, the flashing effect of the squares was created by interpolating between black and white for maximum contrast. Because of the limited frames per second provided by the monitors available for this experiment (60 frames per second) the chosen frequencies were 15, 10 and 6 Hz. These frequencies were easily attained on the display because they are multiples of the monitor's frames per second. When the classification was concluded, the selected square's centre area briefly turned green in colour to indicate the classifier chose that square as selected.

3.1.2.3.2 Signal Conditioning

Before extracting features, both the P300 and the SSVEP signals were filtered digitally. First, a IIR notch filter was applied to suppress the 60 Hz artefact. Then, a FIR bandpass filter from 5 to 30 Hz was applied to attenuate the high and low irrelevant frequencies. Before applying each filter, a 75-sample-mirrored padding was added to each extremity of the signal.

3.1.2.3.2.1 P300 conditioning

Brain signals were recorded from the minimal optimal electrode positions (PO8, PO7, POZ, CPZ) methodologically identified by Speier et al. (2015). Speier at al. compared a four-electrode, six-electrode and 32-electrode configuration, and found no significant difference in ITR between configurations (28.92, 29.94 and 31.90 bits/minute, respectively). The comparison among average accuracies in online trials had no statistical difference (73.21, 69.28 and 67.57%, respectively). Since the performance of the four-electrode configuration was adequate compared to the larger configurations, the four-electrode configuration was chosen for this study so that less time would be needed for the preparation.

After filtering, the signal was down-sampled by a factor of 15. The down-sampled signal constituted the features used to classify the P300 data with the linear discriminant analysis (LDA) classifier. In total each feature sample contained 32 features (eight features per channel, four channels).

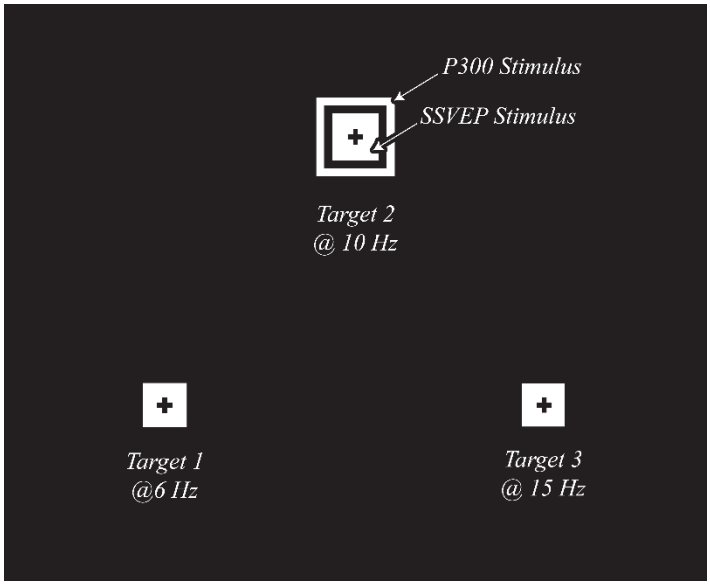

Figure 3.2: Final interface designed to stimulate SSVEP and P300 responses in participants.

3.1.2.3.2.2 SSVEP conditioning

Brain signals were recorded from the electrodes positioned on the occipital region (channels O1, O2) from participants 1 to 3. For participants 4 to 6, the electrodes Pz, T5 and T6 were added (and a sweatband was used to tighten electrodes O1, O2, T5 and T6 to the scalp). For feature extraction, the technique presented by Fan et al. (2015) was used. For each channel, the sum of each target frequency ±0.5 Hz on the power spectral density (i.e. 6, 10 and 15 Hz) and their immediate harmonics (i.e. 12, 20 and 30 Hz) ±0.5 Hz constituted one SSVEP feature sample. In total, each feature sample contained 12 features for group 1, and 30 features for group 2 (three frequencies, plus their harmonics, per channel). Filtered signals were classified using a multiclass LDA classifier.

3.1.2.3.3 Selection functions

The selection functions were only active during online sessions. To define a final output using the classification of each paradigm, the selection functions considered the accuracy of each classifier (calculated in the offline session, during training) and calculated the most likely final answer. A somewhat similar approach was taken by Zuo et al. (2019) where they used a predictive mean calculated with the BLDA (Bayesian Linear Discriminant Analysis) as the classification confidence for the selection between MI and P300. Depending on whether classification confidence lies above or below a threshold, calculated offline via cross-validation, the MI or P300 classification was selected. The general idea of the selector functions in this study was to give points to the paradigms according to their

classification accuracy, rewarding higher accuracies over lower accuracies. Figure 3.3 shows the concept behind the algorithms for each paradigm function. The output of the selection functions was used to provide feedback during the online trials and also to calculate the final accuracy of each paradigm.

The selection functions worked with answer vectors that indicated what the selection score was for each target. For the SSVEP, each vector row corresponded to one target (y_S); for the P300, since it had only target/non-target states (y_P), the position (pos) in which the P300 stimuli appeared corresponded to a vector row. When one of the vector rows surpassed the decision threshold of 3, the corresponding target was selected as the final classification. A timeout was implemented in case the selection function did not reach the minimum decision threshold. In case of timeout, the target with the highest score was selected as the final answer. The timeout was set to 10.5 seconds (21 trials × 0.5s).

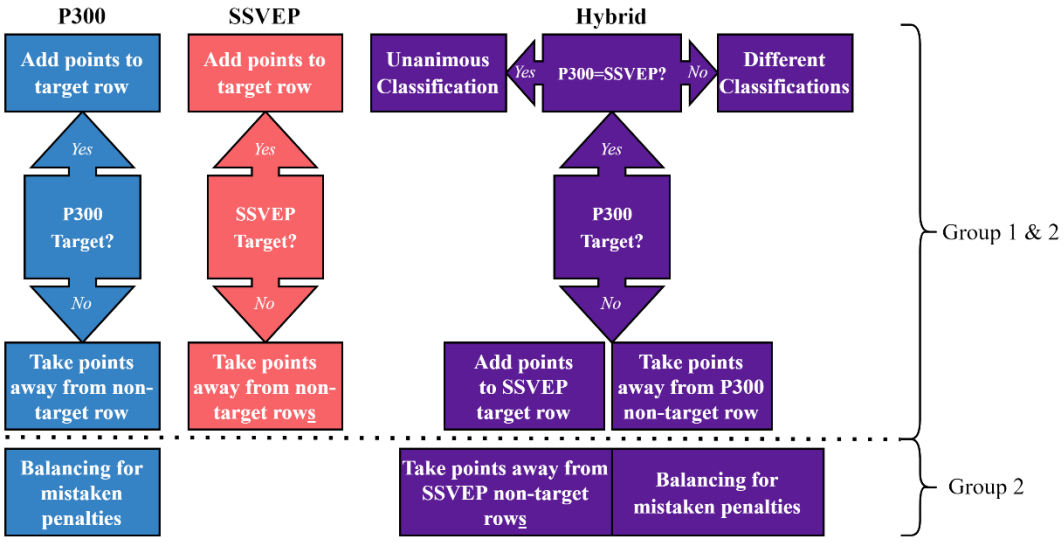

Figure 3.3: General Algorithm for the selection functions

3.1.2.3.3.1 Group 1

For the P300, the answer vector PV was used to score the targets. Each index of PV corresponded to one of the selection scores corresponding to a position pos. The scoring followed the logic of Equation (3.4),

$$PV = \begin{cases} PV_{pos} = PV_{pos} + 2 \times ACC_{P300}, & if\ y_P = 1 \\ PV_{pos} = PV_{pos} - 0.5 \times ACC_{P300}, & if\ y_P = 0 \end{cases} \quad (3.4)$$

where pos was the position of the analyzed P300 stimuli, which equates to the vector's index, y_P was the P300 LDA classification output and the ACC_{P300} was the accuracy calculated during the training session for the P300 classifier. If y_P yielded a target (=1), the index corresponding to that target increases; if non-

target (=0), that index decreased. The weight of +2 for the accuracy was chosen when $y_P = 1$ so that a high accuracy could achieve a value close to the threshold that would only need one or two more trials to confirm the final selection. The weight of -0.5 was chosen if $y_P = 0$ so the decrease could be a penalty that would not penalize a misclassification too much.

For the SSVEP, the answer vector SV was used to score the targets. Each index of SV corresponded to one of the selection score of the classifier's outputs. The scoring followed the logic of Equation (3.5),

$$SV = \begin{cases} SV_{y_S} = SV_{y_S} + 2 \times ACC_{SSVEP} \\ SV_{\overline{y_S}} = SV_{\overline{y_S}} - 0.5 \times ACC_{SSVEP} \end{cases} \quad (3.5)$$

where y_S was the SSVEP LDA classification, $\overline{y_S}$ were the other positions different from the SSVEP LDA classification, and the ACC_{SSVEP} was the accuracy calculated during the training session for the SSVEP classifier. The weights were established following the same logic as in the P300 selector, to reward quicker, and penalize moderately.

The hybrid selection function combined the logic of both selector functions, Equations (3.4) and (3.5). However, instead of $2 \times ACC$, the hybrid selector added 2 points in case of a unanimous answer between the P300 and SSVEP. There were three vectors, one corresponding to the P300 LDA classifier $\boldsymbol{PV}$, one corresponding to the SSVEP LDA classifier $\boldsymbol{SV}$ and one for unanimous answers $\boldsymbol{UV}$, when both classifiers outputted the same target. They had sizes [3x1] in which each of the vector indexes corresponded to the selection score of a target (represented in Figure 3.4). The scoring followed the logic of Equation (3.6),

$$TV = UV + PV + SV = \begin{cases} \begin{cases} UV_{y_S} = UV_{y_S} + 2, & pos = y_S \\ \begin{pmatrix} PV_{pos} = PV_{pos} + ACC_{P300} \\ SV_{y_S} = SV_{y_S} + ACC_{SSVEP} \end{pmatrix}, & pos \neq y_S \end{cases}, & if\ y_P = 1 \\ \begin{pmatrix} PV_{pos} = PV_{pos} - 0.5 \times ACC_{P300} \\ SV_{y_S} = SV_{y_S} + ACC_{SSVEP} \end{pmatrix}, & if\ y_P = 0 \end{cases} \quad (3.6)$$

The logic first defined if the P300 classification was a target or a non-target. If the P300 was classified as a target, the logic verified if the position of the P300 matched the classification of the SSVEP. If the P300 position and the SSVEP matched, then ***UV*** gets points. If they did not match, ***PV*** and ***SV*** get points. If the P300 was classified as a non-target, ***PV*** received a penalty, and ***SV*** increased proportionally to the SSVEP's accuracy. After each trial, all the vectors were summed in the total answer vector ***TV***.

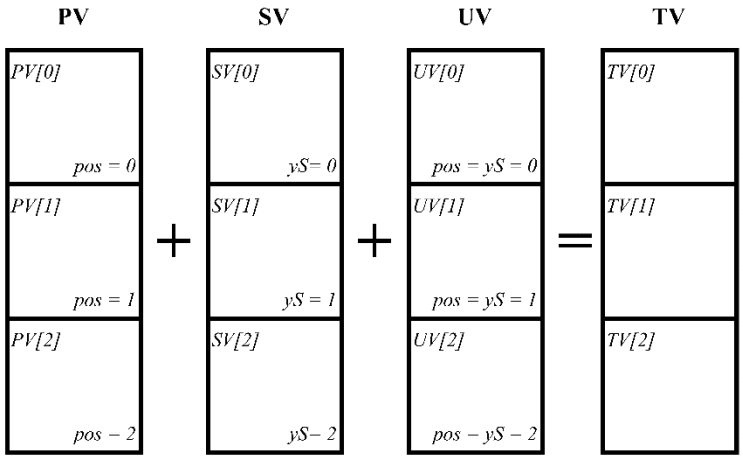

Figure 3.4: Vector scheme for PV. SV, UV and TV

3.1.2.3.3.2 Group 2

For the P300, the logic was similar to Equation (3.4), and is presented in Equation (3.7). The only addition was that when the stimulus was classified as non-target, the other positions different from the analyzed P300 stimuli, increased by half its value (0.25), as a form of balancing for mistaken penalties:

$$PV = \begin{cases} PV_{pos} = PV_{pos} + 2 \times ACC_{P300}, & if\ y_P = 1 \\ \begin{pmatrix} PV_{pos} = PV_{pos} - 0.5 \times ACC_{P300} \\ PV_{\overline{pos}} = PV_{\overline{pos}} + 0.25 \times ACC_{P300} \end{pmatrix}, & if\ y_P = 0 \end{cases} \quad (3.7)$$

where $\overline{pos}$ were the other position indexes.

For the SSVEP, the only modification to Equation (3.5) was a cumulative variable c, presented in Equation (3.8). It was initialized at 1 and increased by 10% for every new classification that matched the previous one, as a form of rewarding classification consistency. The cumulative variable returned to 1 whenever the sequence was broken.

$$SV = \begin{cases} SV_{y_S} = SV_{y_S} + 2 \times ACC_{SSVEP} \times c \\ SV_{\overline{y_S}} = SV_{\overline{y_S}} - 0.5 \times ACC_{SSVEP} \end{cases} \quad (3.8)$$

For the hybrid, the functions were updated to match the selector functions, Equation (3.7) and (3.8), as seen in Equation (3.9). The other changes were: 1) UV received the sum of both classifier accuracies instead of +2, and 2) the weights of penalties and increments were divided by half, as a form of equally sharing the scoring power between both paradigms,

$$TV = UV + PV + SV = \begin{cases} \begin{cases} UV_{y_S} = UV_{y_S} + ACC_{SSVEP} + ACC_{P300}, & pos = y_S \\ \begin{pmatrix} PV_{pos} = PV_{pos} + ACC_{P300} \\ SV_{y_S} = SV_{y_S} + ACC_{SSVEP} \times c \\ SV_{\overline{y_S}} = SV_{\overline{y_S}} - 0.25 \times ACC_{SSVEP} \end{pmatrix}, & pos \neq y_S \end{cases}, & if\ y_P = 1 \\ \begin{pmatrix} PV_{pos} = PV_{pos} - 0.25 \times ACC_{P300} \\ PV_{\overline{pos}} = PV_{\overline{pos}} + 0.125 \times ACC_{P300} \\ SV_{y_S} = SV_{y_S} + ACC_{SSVEP} \times c \\ SV_{\overline{y_S}} = SV_{\overline{y_S}} - 0.25 \times ACC_{SSVEP} \end{pmatrix}, & if\ y_P = 0 \end{cases} \quad (3.9)$$

Figure 3.5 exemplifies what values the vectors would receive under different input conditions in the function implemented for group 1 and 2.

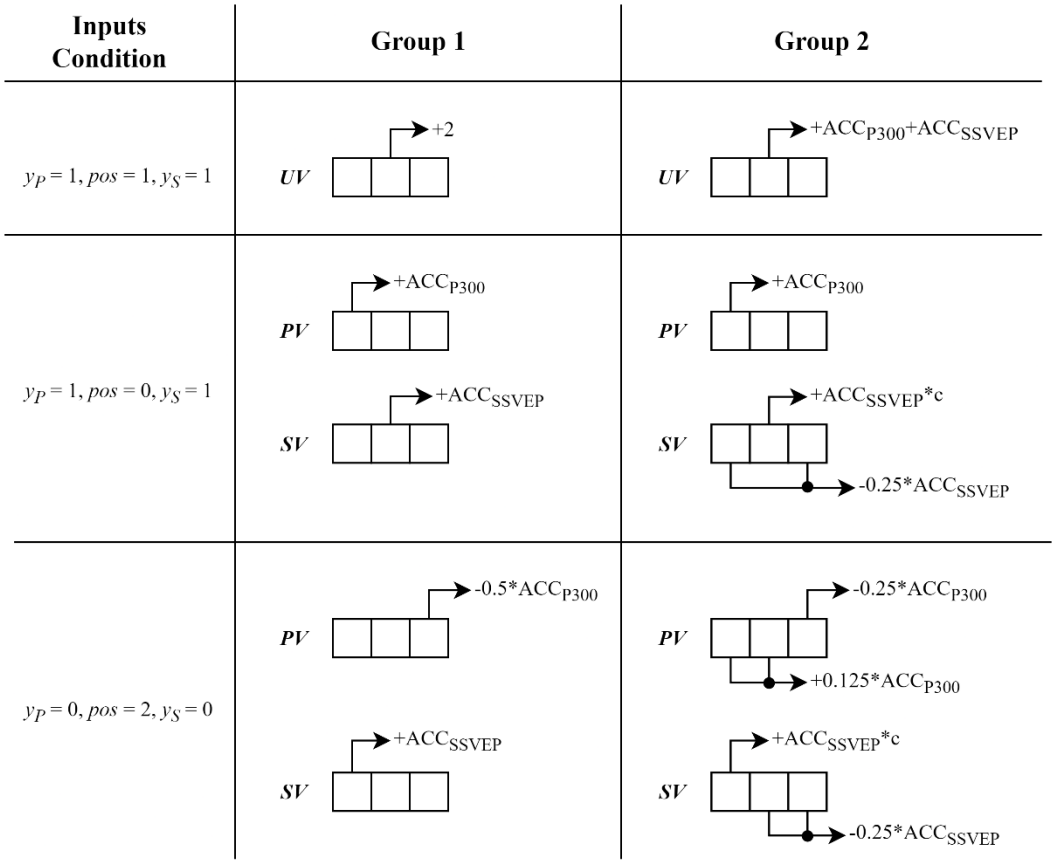

Figure 3.5: Behaviour examples of the hybrid fusion algorithm

3.1.2.4 *Procedure*

In each of the sessions, participants tested one of the single input BCI paradigms (P300 or SSVEP, randomly assigned) or the hybrid BCI. To account for fatigue factors that can cause undesired variability in the results, sessions were scheduled for the same time of the day for each participant. Doing so can standardize the effects of secondary factors (such as hunger, tiredness, distress) on participants' fatigue.

Asking participants to do all the three paradigms allowed the hybrid BCI and the single input BCI to be compared in effectiveness (accuracy) and efficiency (processing time). It also gave participants the opportunity to experience both single input and hybrid BCI so they could give informed advice about their satisfaction with each paradigm.

Before the first session, the purpose of the study was explained to each participant. There were 6 sessions in total over three weeks. Every experimental day, one session took place. The learning effect was attenuated by assigning paradigms to sessions randomly. Sessions were ideally scheduled in a way participants had at least one-day between sessions. Figure 3.6 shows a sample schedule for a participant's sessions.

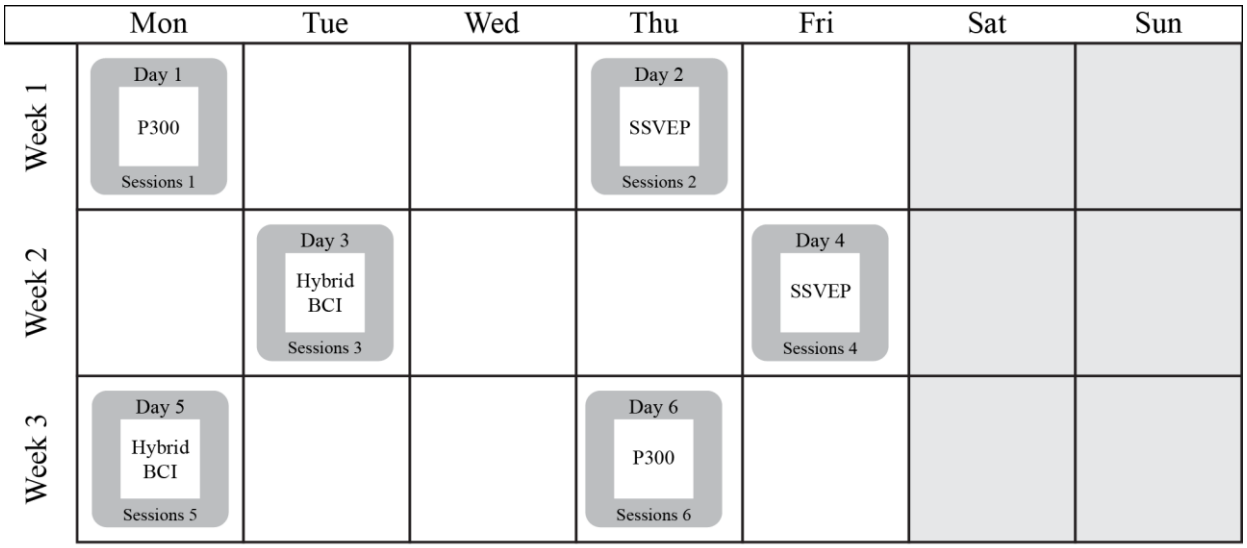

Figure 3.6: Ideal participant schedule.

In each first session with a new paradigm, the participant had an opportunity to practice so they could understand the interface and no results were saved. In subsequent sessions with the same paradigm, practice was offered but only provided if the participant asked for it.

Afterwards, participants underwent five sets of data acquisition, as shown in Figure 3.7. The first set had no selection feedback and served exclusively to train the classifier(s). This set had more runs to allow

more data for the classifier(s) training. The four subsequent sets were online and presented feedback based on the classification.

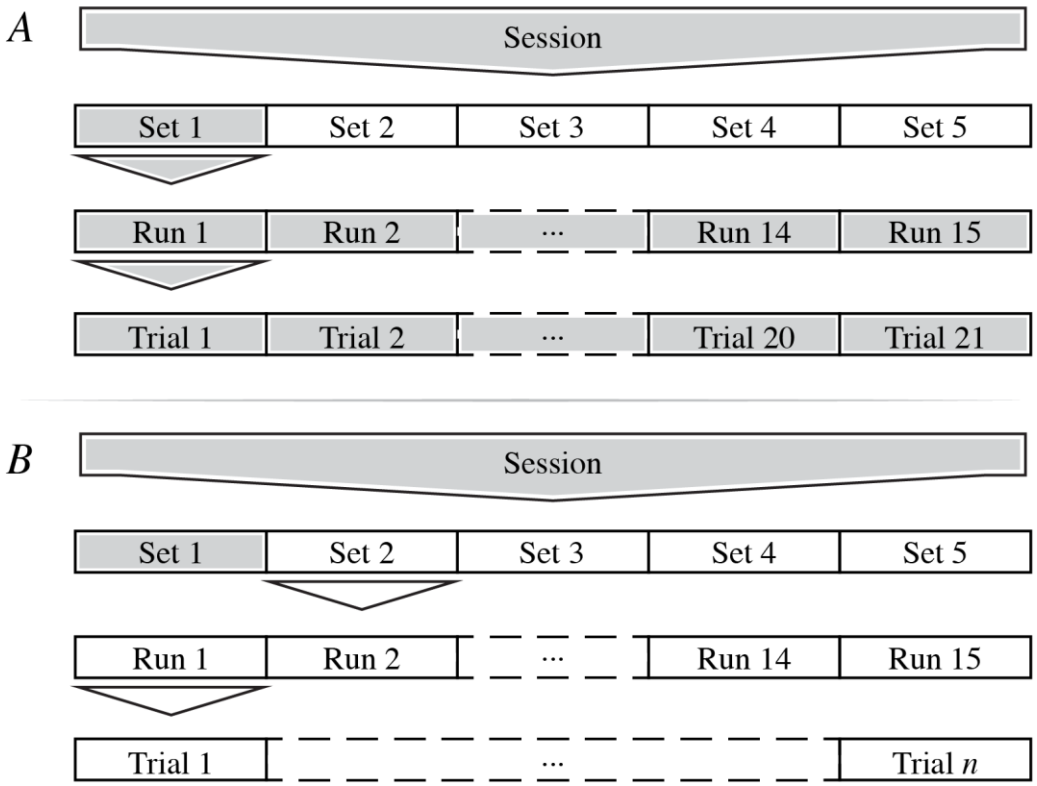

Figure 3.7: A) Offline set of a Session, B) Online set of a Session.

The training set was composed of 15 runs (5 repetition per target × 3 targets), each with a different target order to avoid participants subconsciously attempting to predict cues. Each run was composed of 21 trials (7 trials per target × 3 targets). Considering there were 3s for the cue display, each set took around 3m 25s $(0.5s \times (15\ runs \times 21\ trials) + 3s \times 15\ runs)$. Each training session produced 315 trials for training the classifier (105 for each SSVEP frequency, and 105 targets and 210 non-targets for the P300).

The subsequent online sets had the same number of runs but the number of trials varied depending on the selection functions. A timeout of 10.5s was implemented, as described above, in case the selection functions did not reach the threshold. Only one attempt was allowed per trial.

As mentioned in part 1, for the pure P300 paradigm, the participant was instructed to fixate their gaze on the cued square and count the frame appearances, and for the pure SSVEP paradigm, the participant was instructed to fixate their gaze on the middle cross. For the hBCI, both were combined.

After each session, the participants were asked to fill in a NASA TLX questionnaire regarding the following factors: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (P), Effort (E) and Frustration (F). The questions asked for each factor are:

1. How mentally demanding was the task? (Mental Demand)
2. How physically demanding was the task? (Physical Demand)
3. How hurried or rushed was the pace of the task? (Temporal Demand)
4. How successful were you in accomplishing what you were asked to do? (Performance)
5. How hard did you have to work to accomplish your level of performance? (Effort)
6. How insecure, discouraged, irritated, stressed, and annoyed were you? (Frustration Level)

Each factor was measured through a Likert scale from zero to 10, subdivided by half-points. When all sessions were concluded, a "comparison cards" page was given to each participant to determine a weight to apply to each factor. Each factor was compared against every other factor and the participant chose the one they felt was more important.

If any unforeseen events occurred or any verbal comments were made during the session, a file called "Session_Notes.txt" was created to report those occurrences.

A list with all the steps of the session protocol is attached in Appendix 3. Additional steps to abide by COVID-19 restrictions and approvals to conduct research are in Appendix 4.

3.1.2.5 Data collection and analysis

3.1.2.5.1 Offline sets

To calculate the offline accuracy, a 10-fold cross-validation evaluation was used. The average accuracy will be denominated "offline accuracy" hereafter. There was one offline accuracy for the pure P300 session, one offline accuracy for the pure SSVEP session, and two offline accuracies for the hybrid sessions (the hybrid P300 and the hybrid SSVEP).

3.1.2.5.2 Online sets

During the online sets, every trial was classified and sent to the selector function. Therefore, two classifications were yielded after the online sets, the continuous accuracy and the selector accuracy. The continuous accuracy was calculated based on how many trials were classified correctly. The selector accuracy was calculated based on how many final answers matched the target stimuli.

To calculate the selection time response of the selector function, the number of trials needed for a final answer was counted. Since every trial took half-second, the number of trials multiplied by 0.5 yielded the

selection time response. For each session, 60 selection time responses (15 selection times per set × 4 online sets) were averaged together.

3.1.2.5.3 Statistical analysis

A one-way repeated measures ANOVA was used to compare group mean differences due to repeated measure design occurring in the accuracy and efficiency of paradigms. (Singh et al., 2013).

To examine if there were learning effects between sessions, the values of the first and second sessions of each paradigm were compared. Other comparisons made were between the offline and online accuracies, the pure and hybrid paradigms, between groups, and between average response times. To reject the null hypothesis that the group means were statistically equal, the p-value = 0.05 was used. For the multiple comparisons between offline, continuous and selector accuracies, the Bonferroni correction was used (Bland & Altman, 1995).

3.1.2.5.4 NASA TLX analysis

In total, 15 comparison cards were presented, which was the total amount of weight that was distributed among factors. Since each factor was compared to the other 5 factors, the maximum achievable weight for a factor was 5. A weight value of 0 meant no importance and a value of 5 meant most influential on the final workload. Each factor rating was then multiplied by the resulting weight from the comparison cards. The final workload was calculated by summing all the weighted factors and dividing them by 15. Since each paradigm had two sessions, two final workloads were generated per paradigm. The average of both was considered the average final workload for that paradigm.

The Cohen Kappa inter-reviewer agreement measurement was used to test if the participants consistently agreed in their ranking of the paradigms. Each comparison between participants yielded a score from 0 to 1. As explained by Viera & Garrett (2005), there are six levels of agreement:

- <0 Less than chance
- 0.01 – 0.20 Slight
- 0.21 – 0.40 Fair
- 0.41 – 0.60 Moderate
- 0.61 – 0.80 Substantial
- 0.81 – 0.99 Almost perfect

4 Results

In this section, only results that are statistically significant ($p<0.05$) will be mentioned. More details on the statistical significance of the comparisons will be indicated in the tables.

4.1 Offline results

The average offline accuracies are shown in Figure 4.1 and Table 4.1. Single-paradigm sessions had only one offline accuracy, while the hybrid sessions had two offline accuracies (one for the SSVEP and one for the P300). Figure 4.1A presents the offline accuracies and the standard deviation (SD) for the first and second sessions of each participant.

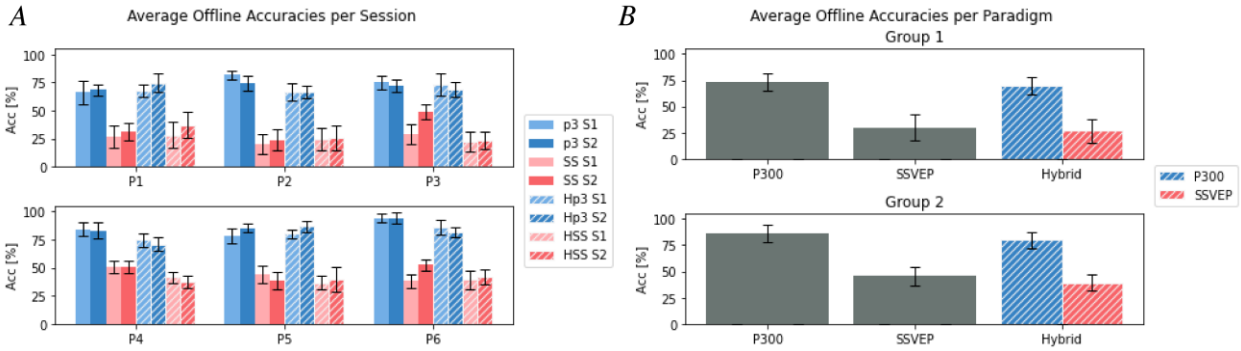

Figure 4.1: Average Offline Accuracies

A) Average offline accuracies per session, B) Offline accuracies averaged per paradigm. The top plots refer to group 1 and the bottom to group 2. The hashed bars represent the accuracies of the paradigms associated with the hybrid system. P: participant, p3: P300, SS: SSVEP, Hp3: Hybrid P300, HSS: Hybrid SSVEP, S: session.

Table 4.1 - Average Offline accuracy comparison between 1st and 2nd session [%±STD]

	P300		SSVEP		Hybrid P300		Hybrid SSVEP	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
P1	66.4±10.5	68.2±5.3	26.4±9.7	30.8±7.6	67.6±5.5	74.6±8.1	28±11.5	37.1±11.4
P2	81.6±3.8**	74.3±6.3**	19.9±9	23.5±9.3	66.6±7.5	66.6±5.9	24.7±10.1	25.4±11.3
P3	74.9±5.8	72.4±5.3	29.2±8.7**	48.9±7**	73.1±9.6	68.6±6.3	22.2±8.9	23.2±7.6
P4	84.4±6.4	82.8±7.2	50.5±5.3	50.5±5.8	74.6±5.7	71.1±6.3	41.6±4.9	37.5±5.9
P5	78.4±6.2	85.4±4.3	43.8±7.8	38.1±7.6	80±4.3*	87±5.1*	36.5±6.3	39.7±10.6
P6	94±3.9	94.3±5.2	38.3±6.1**	52.4±5.5**	85.7±6.5	81.9±4.7	39.4±8.1	41.6±6.8

*Note. 1st and 2nd sessions are statistically different with *: $p<0.05$, **: $p<0.01$*

The results did not show any clear trend when comparing the first and the second session of each system, but some pairs had significant differences. Participant 2 had a better performance on the first P300 session compared to the second (81.6% and 74.3%, respectively). Participants 3 and 6 had more than 14% improvement in accuracy from the first to second SSVEP session. Participant 5 had a 7% improvement on the hybrid P300.

The accuracies averaged across participants for each system is presented in Figure 4.1B. **Comparing group 1 and 2,** all paradigms had an increase in accuracy from group 1 to 2, as shown in Table 4.2. The P300 increased 13.6% (from 72.98 to 86.55) and the SSVEP increased 16% (from 29.78 to 45.61). The hybrid components increased 10.5% on the P300 (from 69.53 to 80.06) and 12.6% on the SSVEP (from 26.77 to 39.36).

Table 4.2 – Average Offline accuracy comparison between group 1 and 2 [%±STD]

	Group 1	Group 2	Acc Diff
P300	72.98±8.24	86.55±8.13	13.57*
SSVEP	29.77±12.75	45.6±8.49	15.83*
H P300	69.53±8	80.06±7.87	10.52*
H SSVEP	26.77±11.49	39.36±7.39	12.59*

*Note. All average results from group 1 and group 2 are statistically different *: $p<0.01$*

4.2 Online results

4.2.1 Continuous classification

During the online sets, the new incoming data were classified continuously by the classifier trained during the offline set. Every session generated four continuous accuracies per paradigm, each corresponding to one online set. The average continuous accuracies per session are displayed in Figure 4.2A. Table 4.3 contains the accuracies and an indication of the statistically significant differences between the first and second session.

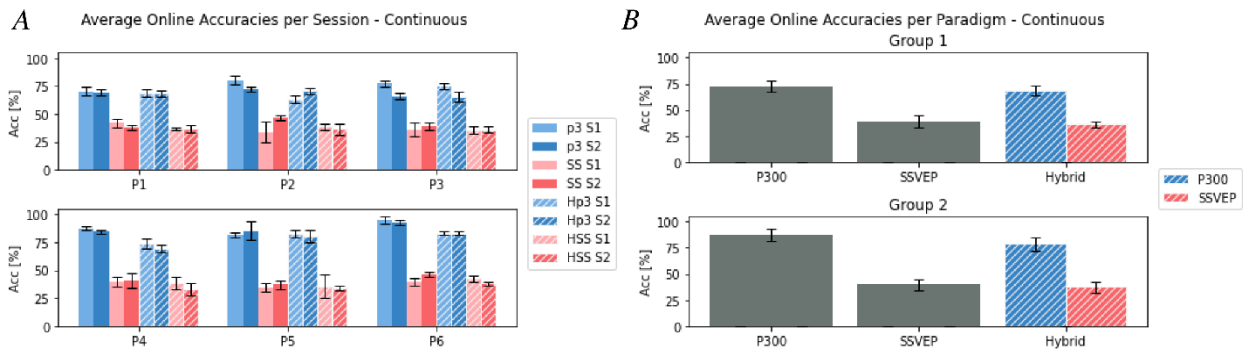

Figure 4.2: Average Online Continuous Accuracies

A) Average Online Continuous Accuracies per Participant, B) Online Continuous Accuracies averaged per Paradigm. The top plots refer to group 1 and the bottom to group 2. The hashed bars represent the accuracies of the paradigms associated with the hybrid system. P: participant, p3: P300, SS: SSVEP, Hp3: Hybrid P300, HSS: Hybrid SSVEP, S: session.

Table 4.3 - Average online continuous accuracy comparison between 1st and 2nd session [%±STD]

	P300		SSVEP		Hybrid P300		Hybrid SSVEP	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
P1	70.1±4.1	69.2±2.8	41.6±4.3	37.5±2.2	69±3.1	68.3±2.9	36.4±1.1	36.2±3.3
P2	80.2±4.2*	72.4±2.4*	33.3±9.4	46.5±2.6	63.2±3.4*	70.5±3*	38.4±3.2	36.4±5.2
P3	77.1±2.8**	65.8±3.1**	36±6.5	38.8±3.6	75.1±2.9**	65.1±4.3**	35.4±3.1	35.6±2.7
P4	87.5±1.8	84.6±1.6	39.4±4.8	40.8±6.5	74.3±4.4	69.1±3.3	38.8±5.7	33.4±5.5
P5	81.4±2.1	85.3±8.4	34.6±3.6	36.9±3.4	82.9±3.4	80.4±5.8	35.6±10.5	34.2±2.4
P6	94.8±3	93.1±2.1	39.5±3.1*	46.1±1.9*	83.2±1.9	83±1.5	42.8±2.8	38.3±1.7

*Note. 1st and 2nd sessions are statistically different with *: p<0.05, **: p<0.01*

Only five paired sessions had significant differences. Participant 2, had a higher accuracy on the first session of the P300 (+7.8%) and higher second session for the hybrid P300 (+7.2%). Similarly, participant 3 had a higher second session for the P300 and the hybrid P300. Participant 6 had an accuracy increment of 6.5% from the first to the second session of the SSVEP.

Figure 4.2B shows the average online continuous accuracy for all participants. The P300 paradigm had an increase in accuracy when comparing group 1 and 2, from 74.5 to 87.8% (+15.3%), respectively. The hybrid P300 component accuracies were also different. Group 1 had 68.5% and group 2 78.8% (+10.3%).

The continuous accuracy progress throughout the online sets can be seen in Figure 4.3. The purpose of graphing this data is to visually examine any trend or abnormality that may have happened within the set. Some noticeable occurrences were: participant 2 in the second session of SSVEP, had an increase from the first session in accuracy for all sets; participant 4 in both sessions with the hybrid P300, was consistently lower than his peers; and participant 6 in both sessions of P300, was consistently above 90% accuracy throughout all of the sets.

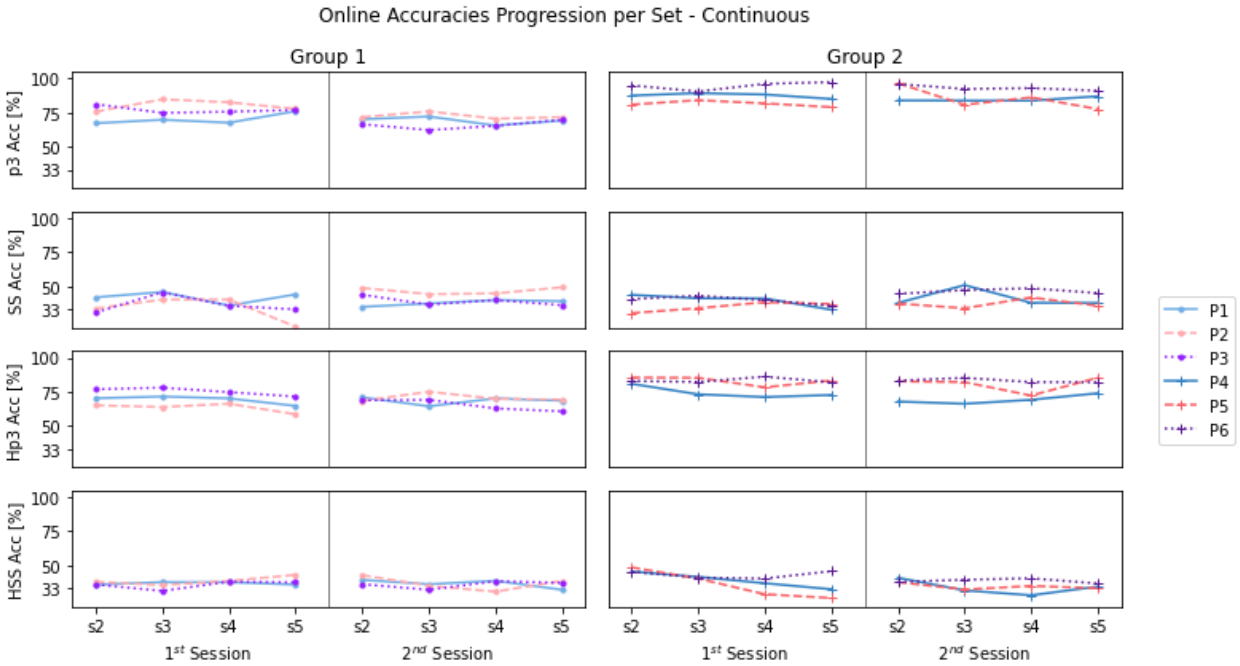

Figure 4.3: Continuous Accuracies of each Set per Group
P: participant, Acc: Accuracy, p3: P300, SS: SSVEP, Hp3: Hybrid P300, HSS: Hybrid SSVEP, s: set.

4.2.2 Selector classification

The selector function received the classifiers' outputs and calculated the most likely final answer, hence, each set had one selector accuracy and because there were four online sets per session, each session had four selector accuracies. For each participant, the averaged selector accuracy across sets is presented in Figure 4.4A and Table 4.4.

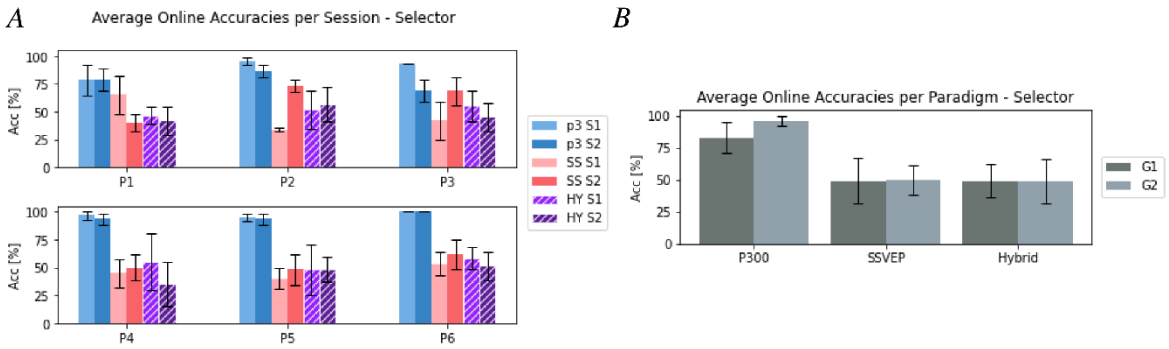

Figure 4.4: Average Online Selector Accuracies

A) Average Online Selector Accuracies per Participant, B) Online Selector Accuracies averaged per Paradigm. The top plots refer to group 1 and the bottom to group 2. The hashed bars represent the accuracies of the paradigms associated with the hybrid system. P: participant, p3: P300, SS: SSVEP, HY: Hybrid.

Most participants presented no statistical difference between the first and second session. In group 1, participant 2 had an increment of 39.6% from the first to the second session of SSVEP and participant 3 had a decrement of 25% from the first to the second session of P300.

The average accuracy per paradigm was calculated based on all set accuracies for all session of all participants in the group (Figure 4.4B). Only the P300 had a significant increase from group 1 to group 2 (+13.1%). The SSVEP and the hybrid had, approximately, the same outcome between groups. The SSVEP and the hybrid also had, approximately, the same outcome when compared in each groups.

Table 4.4 - Average selector accuracy comparison between 1st and 2nd session [%±STD]

	P300		SSVEP		Hybrid	
	1st	2nd	1st	2nd	1st	2nd
P1	78.3±13.7	78.3±10	65±17.5	40±7.7	46.7±7.7	41.7±12.6
P2	95±3.3	86.7±5.4	33.8±1.8**	73.3±5.4**	51.7±17.5	56.7±15.9
P3	93.3±0*	68.3±10*	41.7±17.5	43.3±12.8	55±13.7	45±12.6
P4	96.7±3.8	93.3±5.4	45±12.6	50±11.5	55±25.7	35±19.9
P5	95±3.3	93.3±5.4	40±9.4	48.3±13.7	48.3±22.7	48.3±11.4
P6	100±0	100±0	53.3±10.9	61.7±3.3	58.3±10	51.7±12.6

*Note. 1st and 2nd sessions are statistically different with *: p<0.05, **: p<0.01*

The set accuracies for each selector function are plotted in Figure 4.5. For the P300 sessions, group 2 had the most consistent results. Participant 2 was an outlier in the second session of SSVEP, with higher accuracies than the other two participants in his group. In most sessions with the hybrid, there was a downtrend from one set to the next for most participants in the first and second session (except for participant 6).

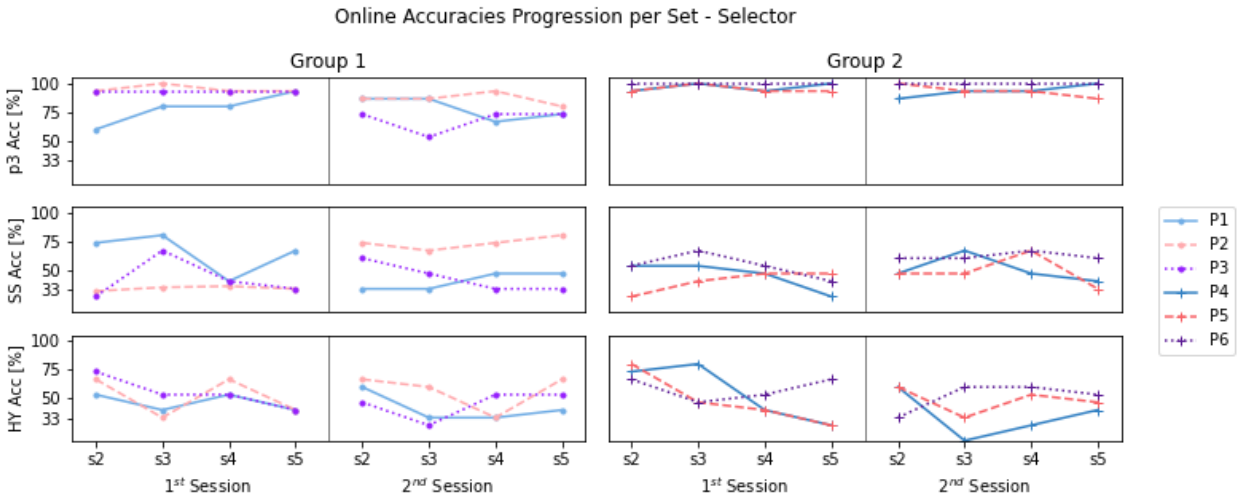

Figure 4.5: Set Selector Accuracies per Group
P: participant, p3: P300, SS: SSVEP, HY: Hybrid, s: set.

4.3 Offline vs Online

Figure 4.6 shows the offline accuracy and the two online accuracies, the continuous and the selector's accuracy, for each group.

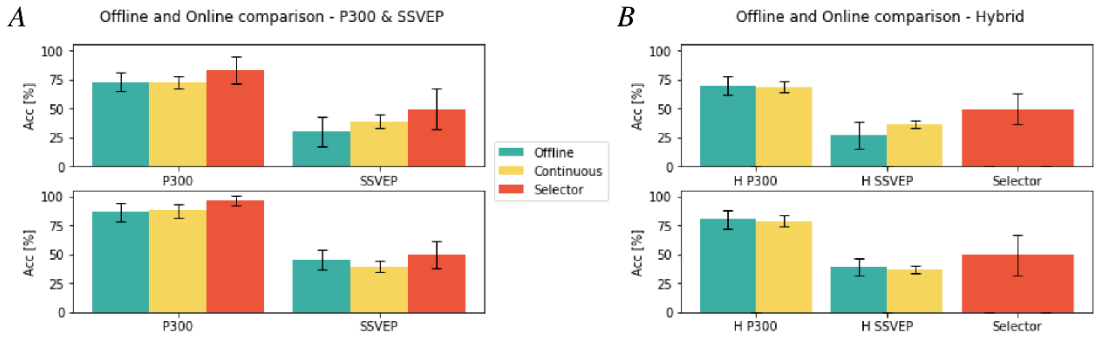

Figure 4.6: Comparison for the Offline, Continuous and Selector average accuracies
A) Compariong the pure SSVEP and P300, B) the hybrid, and the hybrid P300 and SSVEP. The top plots refer to group 1 and the bottom to group 2.

Table 4.5 shows the average accuracy values and their differences. For group 1, all the compared average accuracies were significantly different, except for the accuracies of the P300 offline and continuous, and of the hybrid SSVEP offline and continuous. In group 2, all the compared average accuracies were significant, except for the accuracies of the P300 offline and continuous, of the SSVEP offline and

selector, of the hybrid SSVEP offline and continuous, and the hybrid SSVEP offline and the hybrid selector. For both groups, the selector accuracy was higher for the P300 and the SSVEP. The hybrid selector accuracy was about the average of the hybrid P300 and the hybrid SSVEP.

Table 4.5 – Differences among average accuracies for each paradigm [%]

	Group 1		Group 2	
	Acc	Difference	Acc	Difference
P300 O	72.979	-0.523	86.551	1.247
P300 C	72.456		87.798	
P300 O	72.979	10.351**	86.551	9.838**
P300 S	83.330		96.389	
P300 C	72.456	10.874**	87.798	8.591**
P300 S	83.330		96.389	
SSVEP O	29.775	9.197**	45.608	-6.051**
SSVEP C	38.972		39.557	
SSVEP O	29.775	19.740**	45.608	4.114
SSVEP S	49.515		49.722	
SSVEP C	38.972	10.543**	39.557	10.165**
SSVEP S	49.515		49.722	
H P300 O	69.533	-1.018	80.057	-1.242*
H P300 C	68.515		78.815	
H P300 O	69.533	-20.089**	80.057	-30.613**
Hybrid S	49.444		49.444	
H P300 C	68.515	-19.071**	78.815	-29.371**
Hybrid S	49.444		49.444	
H SSVEP O	26.773	9.642	39.362	-2.186
H SSVEP C	36.415		37.176	
H SSVEP O	26.773	22.671**	39.362	10.082
Hybrid S	49.444		49.444	
H SSVEP C	36.415	13.029**	37.176	12.268**
Hybrid S	49.444		49.444	

*Note. Average accuracy comparisons are statistically different with *: $p<0.0167$, **: $p<0.0033$ with the Bonferroni correction. H: Hybrid, O: offline, C: continuous, S: selector.*

4.4 Efficiency

The time response for each target depended on how many outputs from the classifier were necessary to reach the threshold of the final answer. Therefore, each target had a single time response associated with its accuracy. The time response for each session was averaged for all the 60 targets (15 targets per set × 4 online sets). Figure 4.7A presents the averaged time responses for the first and second sessions of each participant.

Table 4.6 displays all average times for each paradigm. The table is color coded where the green pairs indicate a significant decrease, while the red indicate a significant increase in time. Out of the 18 paired sessions, 8 presented a decrease in response time from the first to second session. Participants 1, 2, 3 and 6 had faster results on the second session of SSVEP. Participant 4 was the only one that did not have any significant change from the first to the second session in any paradigm.

Table 4.6 - Average time response comparison between 1st and 2nd session [s±STD]

	P300		SSVEP		Hybrid	
	1st	2nd	1st	2nd	1st	2nd
P1	8.6±2.8**	7.2±2.7**	7.9±2.4**	5.9±1.8**	7.8±2.9**	4.6±2.3**
P2	4.8±2.6**	7±2.6**	9.9±1**	8.3±2.1**	7.1±2.9	7.1±3
P3	6.6±2.6**	8±2.4**	6.1±2.3**	4.1±1.4**	7.5±2.8**	5.3±2.5**
P4	3.5±1.8	3.7±1.8	3.2±1.3	3.1±1.1	6.3±2.5	5.6±2.1
P5	3.9±2.3*	3.3±1.5*	3.9±1.6**	5.1±1.9**	4.8±1.6	4.6±1.8
P6	2.8±0.9	3±1.1	4.9±1.9**	3±1.1**	4.9±7.5	4.2±3.5

*Note. 1st and 2nd sessions are statistically different with *: p<0.05, **: p<0.01*

Figure 4.7B shows the comparison of the systems' response time between group 1 and 2. Group 2 took less time for all paradigms. The time response for the P300 went from 7.03 to 3.35 s, for the SSVEP from 7.04 to 3.86 s and for the hybrid from 6.57 to 5.07 s.

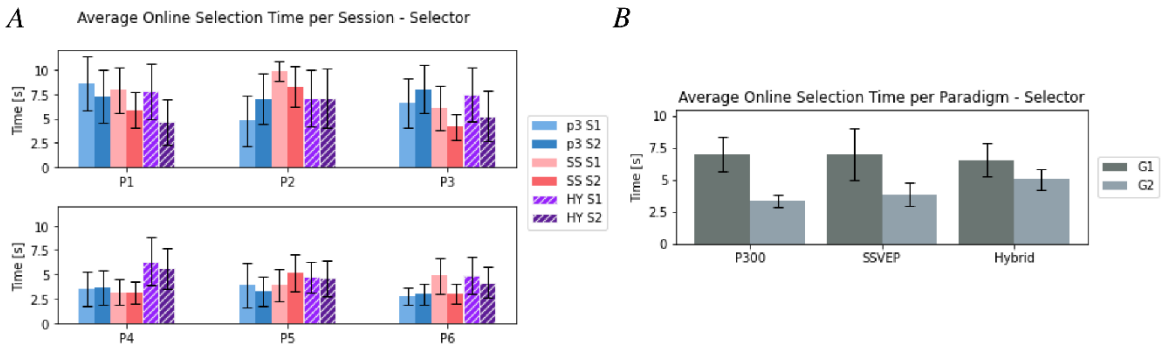

Figure 4.7: Selector Time Response per Group

A) Selector time responses per group, averaged from all sets' time responses. B) Average accuracy across participants for group 1 and 2. The top plot refers to group 1 and the bottom to group 2. P: participant, p3: P300, SS: SSVEP, HY: Hybrid.

Table 4.7 shows the ITR calculated through Equation (3.3) for each paradigm and accuracy (offline, continuous and selector). The number of stimuli per minute was 96 for the offline and 84 for the online (m), considering one stimulus was presented every half-second and 3s between runs and 3s to display the feedback. The number of commands was 3 (N). The best ITR were from the P300 in group 2. The highest was the selector with 111.27 bits/min, then the continuous with 89.07 bits/min, followed by the offline

with 84.56 bits/min. All P300 and hybrid P300 ITR remained above 35 bits/min. All the SSVEP and hybrid SSVEP ITR remained under 7 bits/min. The P300 and the SSVEP had higher ITR for the selector.

Table 4.7 - ITR per paradigm

Type	Group 1		Group 2	
	Acc [%]	ITR [bits/min]	Acc [%]	ITR [bits/min]
P300 O	72.979	45.406	86.551	84.561▪
P300 C	72.456	44.194	87.798	89.071♦
P300 S	83.333	64.535	96.389	111.273†
SSVEP O	29.775	-	45.608	4.475
SSVEP C	38.972	0.966	39.557	1.174
SSVEP S	49.515	6.735	49.722	6.905
H P300 O	69.533	37.763	80.057	63.816
H P300 C	68.515	35.656	78.815	60.297
H SSVEP O	26.772	-	39.362	1.103
H SSVEP C	36.415	0.292	37.176	0.452
Hybrid S	49.444	6.678	49.444	6.678

Note. -: Accuracies below chance level have distorted ITR, so they were not considered, ▪: third best ITR, ♦: second best ITR, †: best ITR, Acc: Accuracy, ITR: Information Transfer Rate, H: Hybrid, O: offline, C: continuous, S: selector.

4.5 NASA TLX questionnaire

Table 4.8 presents the adjustment weights that were applied for each factor attributed by the participants.

Table 4.8 - Adjustment weights for the NASA TLX ratings

	PD	F	TD	E	MD	P
P1	0	5	4	1	2	3
P2	0	2	5	2	4	2
P3	3	0	1	4	2	5
P4	0	2	1	5	3	4
P5	0	1	5	3	4	2
P6	0	4	1	2	4	4
Avg.	0.5	2.33	2.83	2.83	3.17	3.33

Note. PD: Physical Demand, F: Frustration, TD: Temporal Demand, E: Effort , MD: Mental Demand and P: Performance. The color code represents the weights from less important to more important with the colors in the following order: red, orange, yellow, light green, medium green and dark green.

The average final workload for each participant over two sessions of each paradigm is presented in Figure 4.8A. For all participants but participant 3, the P300 had the lowest load. For participants 1, 4, 5 and 6 the highest load was attributed to the hybrid, followed by the SSVEP. Participant 2 had similar ratings for the SSVEP and the hybrid and participant 3 was the only one that rated the SSVEP as the highest load.

An overall workload rating was calculated by averaging all participant's average final workloads for each paradigm. The result is presented in Figure 4.8B. The P300 load was significantly lower than those of the SSVEP and the hybrid.

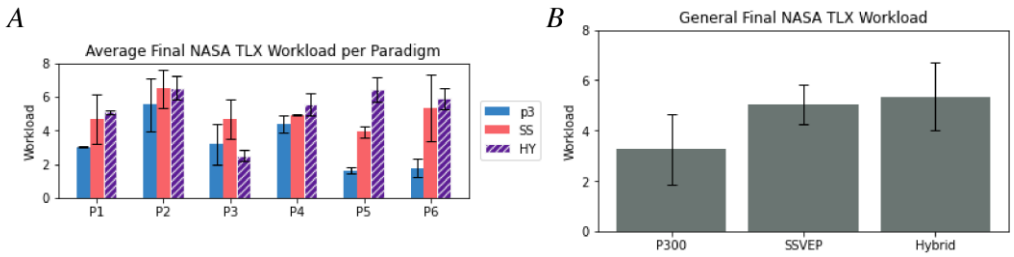

Figure 4.8: Average NASA TLX
A) Average NASA TLX Workload per Participant, B) Average NASA TLX Workload per Paradigm. P: participant, p3: P300, SS: SSVEP, HY: Hybrid.

To compare the raw factors rating and the adjusted factors ratings, Figure 4.9 presents two radar charts showing the effects of the adjustment weights on the average of the average final workload of each paradigm. Low-rated factors had less effect on the final workload, such as Physical Demand (PD), while high-rated factors had more effect on the final load, such as Mental Demand (MD) and Performance (P). Physical Demand had the greatest decrease in rating when weighted. Factors Temporal Demand (TD), Effort (E) and Frustration (F) maintained a similar proportion after the adjustment. Overall, the P300 had the lowest loads for all factors, especially for Frustration. The hybrid had the highest Mental Demand and Temporal Demand, while the SSVEP had the highest Performance load. The SSVEP and the hybrid had similar Effort and Frustration ratings. All paradigms had low Physical Demand.

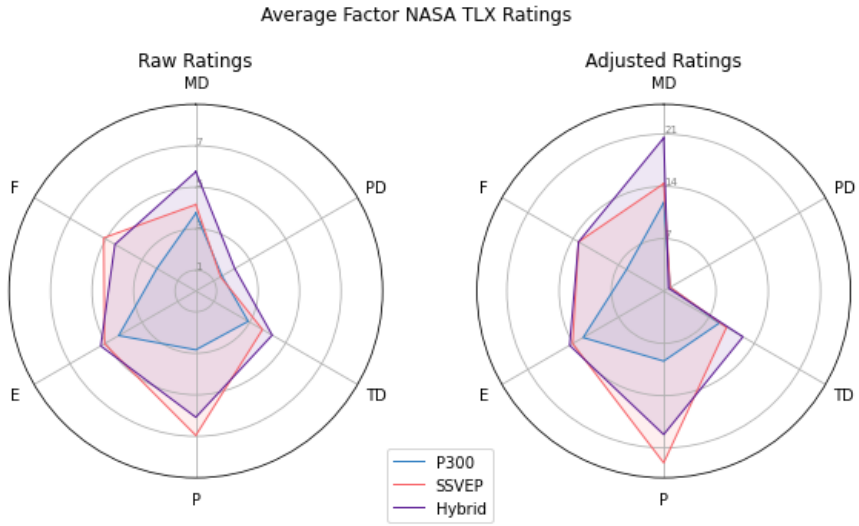

Figure 4.9: Raw and adjusted ratings of the NASA TLX factors
MD: Mental Demand, PD: Physical Demand, TD: Temporal Demand, P: Performance, E: Effort and F: Frustration.

Cohen Kappa agreement scores were calculated among the comparison cards for each factor (Figure 4.10A), among the individual factor ratings (Figure 4.10B) and among the final workload for each paradigm (Figure 4.10C) per participant.

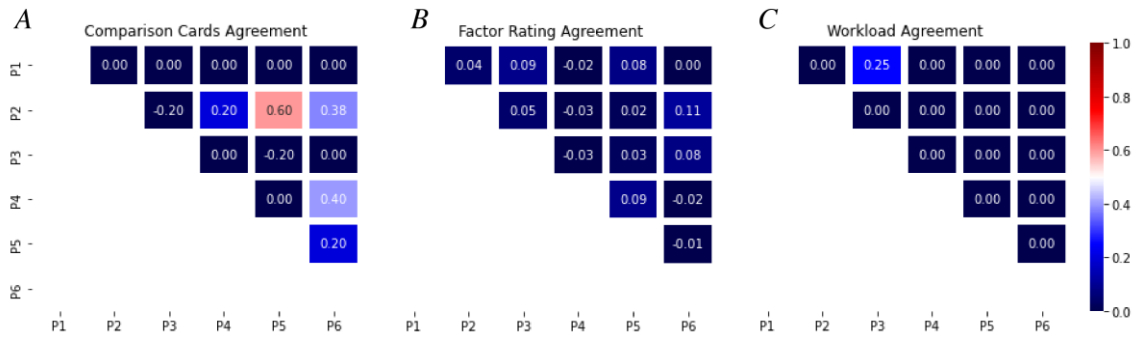

Figure 4.10: Cohen Kappa agreement score between participants

A) Comparison Cards, B) the Individual Factor Ratings and C) the Final Workload. The agreement levels are <0: less than chance, 0.01 to 0.20: slight, 0.21 to 0.40: fair, 0.41 to 0.60: moderate, 0.61 to 0.80: substantial, and 0.81 to 0.99: almost perfect.

Most participants did not agree on the importance of each factor on the comparison cards. Only participants 2, 4, 5 and 6 had some agreement with each other. The individual factor ratings only had slight agreements. For the final workload between paradigms, only P1 and P3 agreed fairly. Therefore, most participants felt differently in regards to the factors' importance, the factor ratings and the final workload for each paradigm.

4.6 Session Notes

Session notes and incidental comments can be seen in Appendix 5.

4.7 Post analysis

After the study, more examination of the data was done to explore the low SSVEP performance, and some tests were done offline to see about improving its performance.

4.7.1 Further examination of the data

4.7.1.1 Hybrid correlation with P300 and SSVEP

The main reason for the hybrid's low accuracy was the low SSVEP performance. To examine this hypothesis, Figure 4.11 shows the scatter plot with the line trend between the SSVEP or P300 component continuous accuracy and the hybrid selector accuracy. The accuracies were set to unit value. The linear trend calculated between both comparisons had a positive slope, indicating a proportional correlation. But Figure 4.11B has a stronger positive slope compared to Figure 4.11A indicating that the SSVEP had higher impact on the hybrid accuracy.

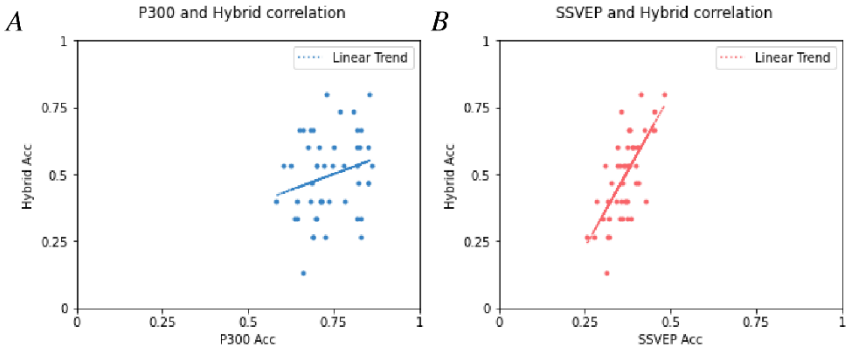

Figure 4.11: Scatter Plot with trend line between Accuracies
A) Continuous hybrid P300 versus hybrid Selector, B) Continuous hybrid SSVEP versus hybrid Selector.

4.7.1.2 Pure paradigms vs. hybrid paradigms

In order to assess if the simultaneous combination of both P300 and SSVEP caused any deleterious effect to the accuracy of the paradigms compared to the pure form, the hybrid and the pure averages were compared and statistically analysed. Table 4.9 shows the accuracy differences ($ACC_{hybrid} - ACC_{pure}$). Most comparisons were statistically significant. The P300 had lower accuracies in all cases when implemented with the SSVEP. On the other hand, the SSVEP, in most cases, had no change when implemented along with the P300. The only exception was in group 1, for the offline accuracy comparison. Almost all the offline comparisons had a decrease in accuracy for the hybrid paradigms.

Table 4.9 - Overall average accuracy difference between the hybrid and pure paradigms [%]

		Offline	Continuous
Group 1	P300	-3.45*	-3.94*
	SSVEP	-3.00	-2.56
Group 2	P300	-6.49**	-8.98**
	SSVEP	-6.25**	-2.38

*Note. Paradigms that are statistically different between the hybrid and pure paradigm with *: $p<0.05$, **: $p<0.01$*

4.7.1.3 Time response and accuracy

The selector functions were designed so that the response time was smaller when the accuracy was higher. Since the selector function is weighted by the offline accuracy, higher accuracies add more points to their selected targets. So in theory, an inverse trend should be seen between time and accuracy. The relationship between the selector accuracy and the time response are shown in Figure 4.12. The only case where time is inversely proportional to accuracy is for the P300. The SSVEP had no relationship and the hybrid had a slight increasing proportional correlation between time and accuracy. The selector accuracies for the P300 were higher than the hybrid, and equivalent for the SSVEP and hybrid (see Figure 4.6). Thus, the expected behavior of inverse proportionality only applies for "high-enough" accuracies, like those of the P300.

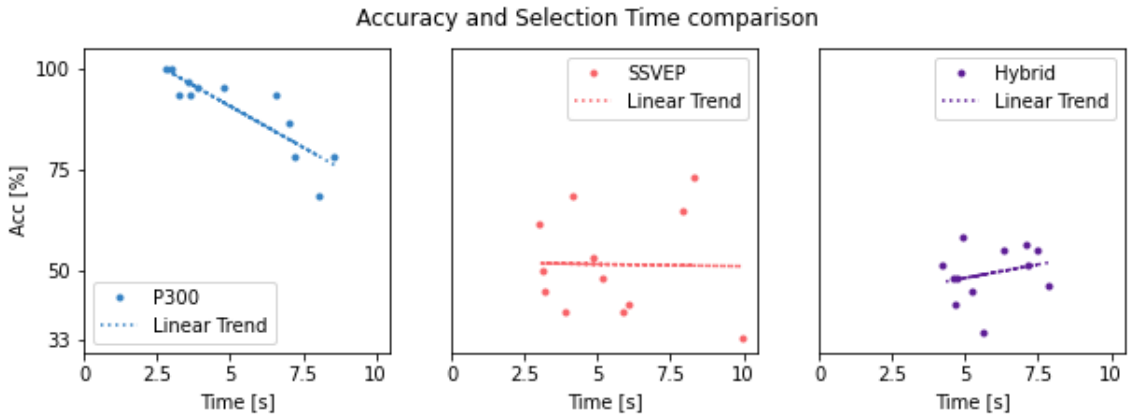

Figure 4.12: Selector Accuracy versus Response Time.

4.7.2 SSVEP improvement tests

First, a progressive sample window was tested, where instead of keeping the window size to 0.5 seconds, the samples were compounded over time (i.e., the first window had only the first 0.5s sample; the second window had the first and second 0.5s samples; the 3rd window had the first, second and 3rd 0.5s samples, and so on). Thus, the window size was 0.5s, and progressed to 10.5 seconds. This technique allows the signals still to be read every half-second, and is expected to improve feature quality, since it has an increasingly larger and redundant window size from which to extract the frequency information. All the other parameters were kept the same, i.e., the SSVEP conditioning was the same for each new window, the features were input to the same classifier, the 10-fold cross-validation method was used to assess the offline accuracies and the same electrodes were used for each group. The final accuracy from the progressive sample window are shown in Table 4.10 along with the offline accuracies for the SSVEP for all participants.

Although there was a consistent improvement offline, due to the increasing window size at every incoming sample, the frequency-domain conversion also had increasing computational cost, making the processing slower. When this implementation was attempted online, the display could not flash at the intended stimuli frequency, causing the brain response to be unrecognizable by the classifier during the online sets. Therefore, this technique could not be implemented successfully in online trials.

Table 4.10 - Comparison of offline SSVEP accuracies for progressive and static window size

	Session	Static half-second window [%]	Progressive Sample window [%]
P01	1st	26.4±9.7	80.3±6.9
P01	2nd	30.8±7.6	75.2±5.6
P02	1st	19.9±9	73.3±6.2
P02	2nd	23.5±9.3	84.4±8.9
P03	1st	29.2±8.7	85.4±7.9
P03	2nd	48.9±7	78.3±9.5
P04	1st	50.5±5.3	85.4±5.8
P04	2nd	50.5±5.8	88.9±8.6
P05	1st	43.8±7.8	86.7±7.9
P05	2nd	38.1±7.6	84.5±5.4
P06	1st	38.3±6.1	86.3±5.4
P06	2nd	52.4±5.5	86.3±4.3

Secondly, the Filter Bank method was tested for the 0.5 window (Chen et al., 2015). Similar to what was done by Chen et al., each sample underwent 10 sub-band filters, ranging from 5 to 105 Hz, to accommodate for harmonics. The original SSVEP conditioning function was modified to only contain the notch filter (since each sub-band of the filter bank was the new band-pass filter for the signal). The chosen filtering method for the filter bank was the sub-band method M3, since it had the best results in Chen et al.. Each sub-band ranged from a lower frequency, $n \times 5$, up to 105 Hz with an extra 2 Hz on each sub-band extreme (i.e., [3-107] Hz, [13-107] Hz,… [93-107] Hz). Then, the correlation coefficient of a CCA filter, compared to the paired sine-cosine waves of each one of the target frequencies and its first and second harmonics, were summed for each frequency. The summation was weighted to compensate for the magnitude decay of higher harmonics (the same optimal parameters used by Chen et al. were used). The final feature vector contained the summed weighted coefficients for each target frequency. To choose the final answer, the maximum value from the feature vector was chosen. Chen et al. had good results using 1.25 seconds at 1000 Hz down-sampled to 250 Hz, however, their results using a 0.5 second window at 250 Hz were around 33.33%, as in this study.

5 Discussion

In this study, a hybrid BCI using P300 and SSVEP was tested by six participants over six sessions (two sessions per paradigm). Participants from group 2 (participants 4 to 6) tested a slightly modified system than group 1 (participants 1 to 3). Each paradigm was measured offline and online with respect to accuracy (offline, online continuous, and online selector), selection time and information transfer rate, and participant satisfaction. The hybrid accuracy, selection time and satisfaction were not as good as the pure P300, but they were better than pure SSVEP. The lower accuracy of the hybrid was a consequence of the lower accuracy of the SSVEP.

5.1 Accuracy trend

Overall, when analyzing the accuracies, there were no trends when comparing the first and second sessions for any accuracy type. Thus, we can infer that there were no learning effects between sessions. Events that happened, listed in Appendix 5 table, can be correlated to less accurate results. For example, in sessions where participants were sleepy (see Appendix 5) , the accuracy tended to be lower: participant 1's first P300 session showed lower accuracies in the first sets; participant 2 in his first SSVEP session and his second P300 session had a lower accuracy; participant 3 on both his second sessions of P300 and SSVEP had lower accuracy; participant 4 had an accuracy low at the set 3 of his second hybrid session; participant 5, for the hybrid sessions had progressively lower frequencies. On participant 5's first SSVEP session, sudden noises started coming from the hallway from set 2 to 5. His accuracy improved after he wore his earbuds to attenuate noise.

5.2 Group differences

The offline accuracy was higher for all paradigms in the second group (see Figure 4.6), which may be due to the changes made to the system. For the SSVEP, adding electrodes likely added information redundancy and improved the quality of the feature vectors fed to the classifier. Also, the sweatband could have improved the contact of the electrodes against the scalp, reducing noise, for both the SSVEP and the P300. It is also possible that the different room for group 2 helped participants to concentrate better and see more contrast on the computer display, since it was a darker room.

For the continuous and selector accuracies, only the P300 paradigms improved substantially from group 1 to group 2, (+15.3% for the continuous and +13.1 for the selector, see Figure 4.6). The hybrid P300 improved only for the continuous accuracy (+10.3%). This indicates that even with the adjustments made to the SSVEP, there was no accuracy improvement for the SSVEP. Most likely, the issue lies with the small window size that was attempted. Although it made theoretical sense from a signal-processing standpoint because of the frequency resolution (calculated with Equation (3.1)), it appears that half-

second did not have enough information to do the classification accurately using the described method. On the other hand, the P300 selector had an increased accuracy thanks to the modified selector function, which compensated for possible mistaken penalties.

5.3 Why was the hybrid less accurate?

It was hypothesized that the hybrid would have a higher accuracy and lower response time than both pure paradigms. The results showed that the hybrid had an accuracy similar to the average of the P300 and SSVEP accuracies or similar to the SSVEP accuracy. The hybrid's low accuracy may result from some factors discussed in the following.

5.3.1 Low SSVEP accuracy

Since the system was designed to combine both paradigms, the consistent SSVEP misclassifications caused the hybrid selector's low accuracy. Figure 4.12 shows that the SSVEP accuracy had a stronger correlation with the hybrid accuracy than the P300. That might indicate that unless the SSVEP achieves a higher accuracy, the high performance of the P300 will not be enough to improve the hybrid accuracy.

To examine if the selector functions were the cause of the low accuracies, see Table 5.1, which summarizes the results of Figure 4.6. For each of the P300, SSVEP, hybrid P300 and hybrid SSVEP, the type of accuracy (offline, continuous, or selector) was examined to see which had the highest accuracy, which were tied for the highest, and which had the lower accuracies. The table shows that the selector functions were effective in raising the continuous accuracy for the pure P300 and the pure SSVEP and for the hybrid SSVEP. Thus, although the selector functions appear to work as intended on the pure paradigms, the hybrid selector could be improved by prioritizing the higher accuracy between the hybrid P300 and the hybrid SSVEP.

Considering the sets where the hybrid SSVEP performed better (e.g. the second sets of the first session of participants 4, 5 and 6, which were 45.2%, 48.3% and 45%, respectively, see Figure 4.3), it is possible to see that the hybrid selector had a higher accuracy than the hybrid selector average (73.3%, 80% and 66.7%, respectively, see Figure 4.5). SSVEP accuracies lower than 45% tended to penalize the hybrid selector's accuracy, even if the P300 portion had accuracies close to 75%.

Table 5.1 - Highest accuracies comparison summary

	P300			SSVEP			Hybrid (P300)			Hybrid (SSVEP)		
	O	C	S	O	C	S	O	C	S	O	C	S
G1	□	□	☑	□	□	☑	▣	▣	□	□	□	☑
G2	□	☐	☑	▣	□	▣	▣	▣	□	□	□	☑

Note. The columns marked with ☑ indicate the highest accuracy, ▣ indicate accuracies tied for the highest, and ☐ indicate the lower accuracies. O = offline, C = continuous, S = selector

5.3.2 Accuracy reduction when combining paradigms

Table 4.9 shows that when the P300 and the SSVEP paradigms were combined, both offline and continuous accuracies of the P300 dropped when comparing hybrid and pure P300. On the other hand, only in group 2, during the offline session, was there a significant decrease from the pure SSVEP to the hybrid SSVEP.

The drop in P300 accuracy suggests that the P300 might have been harder to identify in the hybrid paradigm. As some participants suggested in their comments in Appendix 5, it was harder to count the P300 frame flashes while looking at the SSVEP flashing. The SSVEP flashing might have been distracting and made the participants miss one or more P300 stimuli.

On the other hand, even with the added P300 frame appearances, it might be possible to assume that the SSVEP was not affected as much by the P300. Finding a different way to display the signals simultaneously might ease the multitasking load that was perceived in the presented configuration.

5.3.3 Higher Mental demand

When the participants tested the hybrid system for the first time, some of them verbally stated it was more overwhelming than the others (see Appendix 5). The hybrid paradigm had the highest Mental Demand among the three paradigms (see Figure 4.9). Only two participants felt sleepy during the hybrid sessions, possibly because the hybrid paradigm required multitasking. However, it still demanded a lot from the participants' mental capacity. This increased participants fatigue during the session.

5.4 NASA TLX

Analyzing the NASA TLX results, we see that the P300 had the overall lowest workload and the hybrid had the highest. The hybrid having the highest workload could be expected since participants attributed the Performance and Mental Demand a higher importance than the other factors when weighting for the final workload and the hybrid had the highest score for both, as discussed above.

We also see that the SSVEP and the hybrid had similar workloads, which might indicate that the perceived workload of the SSVEP was transferred to the hybrid. We can see in Figure 4.9 that the Frustration, Effort and Physical Demand were perceived as equal for both paradigms. The pure SSVEP had a high Performance score (high scores indicate failed performance, and low scores indicate perfect performance), likely due to the lack of success the participants had with the SSVEP. The same lack of success was also felt in the hybrid. The inclusion of the P300 along with the SSVEP affected the Mental Demand, but also the perceived Temporal Demand. Since a there was more stimulation with the hybrid, participants may have felt more rushed.

5.5 Limitations and future work

One limitation was that the 15Hz frequency used was within the seizure-inducing frequency range. Some studies have shown that frequencies between 12 and 25 Hz have more potential to induce seizures (Fisher et al., 2005; Okudan & Özkara, 2018). When implementing this system in the future, it would be important to avoid those frequencies.

The poor accuracy of the SSVEP was a limitation in trying to test the hypothesis that a hybrid would be better than either SSVEP or P300 alone. The poor SSVEP accuracy made the hybrid accuracy poor. Since a smaller static window did not work well enough to obtain a high accuracy for the SSVEP, future projects should use a larger window, but still keep the 0.5s window for the P300. This would make the minimum time selection be the required window length of the SSVEP. There would be a tradeoff, of slower time to select a target, for a higher final accuracy. Chen et al. (2015) have shown that it is possible to get window sizes of 1.25s with offline accuracies close to 90%.

Additionally, when testing the progressive sample window, instead of combining all the windows up to 10.5, there might be a possibility of finding a window length that does not require excessive computational power but that also increases the accuracy. Tests can be conducted offline to find the minimum window size that would provide an increment in accuracy and then it can be tested online to evaluate if the stimuli continue to be stable.

In future projects, the selection function for the hybrid could be modified so that it prioritizes the higher performing paradigm (using logarithmic functions, instead of a linear one, for example). Although it would not solve the problem of the low accuracy of the SSVEP, this could be an extra mechanism to prevent an accuracy drop in the event that one of the paradigms does not perform properly.

The hybrid combination of these two steady-state evoked potentials had low accuracy and a high cognitive load for these participants. An alternative could be to combine the P300 frame with a different

style of steady-state evoked potential. For example, the motion visual evoked potential (mVEP) uses flashing that emulates movement instead of flashing that is static (Guo et al., 2008; Ma et al., 2017). Depending on the simulated movement direction or speed, the brain has counter-lateral responses, which can help detect the participant's desired selection (Punsawad & Wongsawat, 2017).

Finally, to attempt to improve the hybrid accuracy, using a single feature set for both paradigms and a feature selector can also help improve the accuracy. Different strategies to combine both paradigms can change how the information is read by the classifier. Since all data would be unified, only one classifier would be needed and the selector functions would not need to be implemented.

6 Conclusion

In this study a hybrid brain-computer interface was developed using SSVEP and P300 simultaneously. The display showed flashing squares at different frequencies that evoked the SSVEP responses, and a frame that appeared pseudo-randomly around each square evoked the P300 response. The signals were collected for 0.5s per trial, conditioned and then classified by two LDA classifiers. In part 1, eight volunteers tested different parts of the system, and the system and experiment were adapted based on the volunteer's opinions.

The results of part 1 showed that one session per day was the maximum that participants could undergo before experiencing mental fatigue. The colours of the squares were changed, since participants said green and red "blurred their view". Finally, a cross was added to the squares, and the square sizes were changed to improve the simultaneous P300 and SSVEP visualization.

In part 2, six male participants tested the hybrid and the pure paradigms. All underwent six sessions, two per paradigm (SSVEP, P300, and hybrid). The offline, online continuous and online selector accuracies were collected. The first three participants (group 1) tested the system resulting from Part 1. The last three participants (group 2) tested a system modified to improve accuracy and response time.

The results of part 2 showed that the hybrid had 49.44% of accuracy, which was not higher than the pure paradigms, as expected. The overall best paradigm was the P300, getting up to 100% accuracy online. Its average accuracy was 83.3% for group 1 and 96.4% for group 2 with an ITR of 64.53 and 111.27 bits/min, respectively. The SSVEP average accuracy was 49.5% for group 1 and 49.7% for group 2, with an ITR of 6.73 and 6.9 bits/min, respectively. The hybrid performed similarly for both group 1 and 2 with 49.4% accuracy and 6.8 bits/min. The ITR was lower for both the SSVEP and hybrid, compared to the P300 because of the lower accuracies.

The changes made in the system from group 1 to 2, improved the response times for all systems. The accuracy and ITR also improved for the P300, but had no significant impact for the SSVEP or the hybrid.

Workload was measured to see how participants felt about using the system. Their ratings varied on each of the factors and they had different opinions about which factor was the most important. The averages of participant's adjustment weights indicate that the most important factor was Performance, followed by Mental Demand, followed by Effort and Temporal Demand. The lowest overall workload was for the P300 with 3.3, followed by the SSVEP with 5 and the hybrid with 5.4.

The overall results allow us to assume that, although it made theoretical sense, the 0.5s window size restricted the amount of information for the SSVEP classifier, making its performance low and directly impacting the hybrid's accuracy. Even though the selector functions were programmed to consider the individual accuracies of the hybrid SSVEP and hybrid P300, the SSVEP had a greater impact on the hybrid's final accuracy than the P300. Several options for improving the system in the future were suggested.

References

Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., & Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *Journal of Neural Engineering*, *16*(1), 011001. https://doi.org/10.1088/1741-2552/aaf12e

Albrecht, J. S., Bubenzer-Busch, S., Gallien, A., Knospe, E. L., Gaber, T. J., & Zepf, F. D. (2017). Effects of a structured 20-session slow-cortical-potential-based neurofeedback program on attentional performance in children and adolescents with attention-deficit hyperactivity disorder: Retrospective analysis of an open-label pilot-approach and 6-month follow-up. *Neuropsychiatric Disease and Treatment*, *13*, 667–683. https://doi.org/10.2147/NDT.S119694

Anaconda—Individual Edition. (n.d.). Anaconda. Retrieved February 13, 2021, from https://www.anaconda.com/products/individual

Bareš, M., Nestrašil, I., & Rektor, I. (2007). The effect of response type (motor output versus mental counting) on the intracerebral distribution of the slow cortical potentials in an externally cued (CNV) paradigm. *Brain Research Bulletin*, *71*(4), 428–435. https://doi.org/10.1016/j.brainresbull.2006.10.012

Batres-Mendoza, P., Ibarra-Manzano, M. A., Guerra-Hernandez, E. I., Almanza-Ojeda, D. L., Montoro-Sanjose, C. R., Romero-Troncoso, R. J., & Rostro-Gonzalez, H. (2017). Improving EEG-Based Motor Imagery Classification for Real-Time Applications Using the QSA Method. *Computational Intelligence and Neuroscience*, *2017*(9817305), 1–16. https://doi.org/10.1155/2017/9817305

Bauer, G., Gerstenbrand, F., & Rumpl, E. (1979). Varieties of the locked-in syndrome. *Journal of Neurology*, *221*(2), 77–91. https://doi.org/10.1007/BF00313105

Blain-Moraes, S., Schaff, R., Gruis, K. L., Huggins, J. E., & Wren, P. A. (2012). Barriers to and mediators of brain–computer interface user acceptance: Focus group findings. *Ergonomics*, *55*(5), 516–525. https://doi.org/10.1080/00140139.2012.661082

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ : British Medical Journal*, *310*(6973), 170.

Borhani, S., Kilmarx, J., Saffo, D., Ng, L., Abiri, R., & Zhao, X. (2019). Optimizing Prediction Model for a Noninvasive Brain–Computer Interface Platform Using Channel Selection, Classification, and Regression. *IEEE Journal of Biomedical and Health Informatics*, *23*(6), 2475–2482. https://doi.org/10.1109/JBHI.2019.2892379

Brodmann, K. (1909). Comparative Localization Study of the Brain According to the Principles of Cellular Structures. *JA Barth Verlag*.

Carelli, L., Solca, F., Faini, A., Meriggi, P., Sangalli, D., Cipresso, P., Riva, G., Ticozzi, N., Ciammola, A., Silani, V., & Poletti, B. (2017). Brain-Computer Interface for Clinical Purposes: Cognitive Assessment and Rehabilitation. *BioMed Research International*. https://doi.org/10.1155/2017/1695290

Carmona, L., Diez, P. F., Laciar, E., & Mut, V. (2020). Multisensory stimulation and EEG recording below the hair-line: A new paradigm on brain computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 1–1. https://doi.org/10.1109/TNSRE.2020.2979684

Cecotti, H., Phlypo, R., Rivet, B., Congedo, M., Maby, E., & Mattout, J. (2010). Impact of the time segment analysis for P300 detection with spatial filtering. *ISABEL 2010 - 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies*, 5 pages. https://hal.archives-ouvertes.fr/hal-00536275

Chen, X., Wang, Y., Gao, S., Jung, T.-P., & Gao, X. (2015). Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. *Journal of Neural Engineering*, *12*(4), 046008. https://doi.org/10.1088/1741-2560/12/4/046008

Cho, H., Ahn, M., Ahn, S., Kwon, M., & Jun, S. C. (2017). EEG datasets for motor imagery brain–computer interface. *GigaScience*, *6*(7). https://doi.org/10.1093/gigascience/gix034

Choi, J. W., Rho, E., Huh, S., & Jo, S. (2018). An EOG/EEG-Based Hybrid Brain-Computer Interface for Chess. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 129–134. https://doi.org/10.1109/SMC.2018.00033

Cincotti, F., Mattia, D., Aloise, F., Bufalari, S., Schalk, G., Oriolo, G., Cherubini, A., Marciani, M. G., & Babiloni, F. (2008). Non-invasive brain–computer interface system: Towards its application as assistive technology. *Brain Research Bulletin*, *75*(6), 796–803. https://doi.org/10.1016/j.brainresbull.2008.01.007

Cioni, G., Guzzetta, A., & D'Acunto, G. (2012). Cerebral Plasticity and Functional Reorganization in Children with Congenital Brain Lesions. In G. Buonocore, R. Bracci, & M. Weindling (Eds.), *Neonatology: A Practical Approach to Neonatal Diseases* (pp. 145–149). Springer Milan. https://doi.org/10.1007/978-88-470-1405-3_21

Collinger, J. L., Boninger, M. L., Bruns, T. M., Curley, K., Wang, W., & Weber, D. J. (2013). Functional priorities, assistive technology, and brain-computer interfaces after spinal cord injury. *The Journal of Rehabilitation Research and Development*, *50*(2), 145. https://doi.org/10.1682/JRRD.2011.11.0213

Conda environments. (n.d.). Retrieved February 13, 2021, from https://docs.conda.io/projects/conda/en/latest/user-guide/concepts/environments.html

Cook, A. M., & Polgar, J. M. (2014). *Assistive technologies: Principles and practice*. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1151635

Diniz, P. S. R., Silva, E. A. B. da, & Netto, S. L. (2002). *Digital Signal Processing: System Analysis and Design*. Cambridge University Press.

Donchin, E., Spencer, K. M., & Wijesinghe, R. (2000). The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 174–179. https://doi.org/10.1109/86.847808

Dovgialo, M., Chabuda, A., Duszyk, A., Zieleniewska, M., Pietrzak, M., Różański, P., & Durka, P. (2018). Assessment of Statistically Significant Command-Following in Pediatric Patients with Disorders of Consciousness, Based on Visual, Auditory and Tactile Event-Related Potentials. *International Journal of Neural Systems*, *29*(03), 1850048. https://doi.org/10.1142/S012906571850048X

Ehlers, J., Valbuena, D., Stiller, A., & Gräser, A. (2012). Age-Specific Mechanisms in an SSVEP-Based BCI Scenario: Evidences from Spontaneous Rhythms and Neuronal Oscillators. *Computational Intelligence and Neuroscience*, 379. https://doi.org/10.1155/2012/967305

Fan, X., Bi, L., Teng, T., Ding, H., & Liu, Y. (2015). A Brain-Computer Interface-Based Vehicle Destination Selection System Using P300 and SSVEP Signals. *Ieee Transactions on Intelligent Transportation Systems*, *16*(1), 274–283. https://doi.org/10.1109/TITS.2014.2330000

Faress, A., & Chau, T. (2013). Towards a multimodal brain–computer interface: Combining fNIRS and fTCD measurements to enable higher classification accuracy. *NeuroImage*, *77*, 186–194. https://doi.org/10.1016/j.neuroimage.2013.03.028

Fernández, E., Greger, B., House, P. A., Aranda, I., Botella, C., Albisua, J., Soto-Sánchez, C., Alfaro, A., & Normann, R. A. (2014). Acute human brain responses to intracortical microelectrode arrays: Challenges and future prospects. *Frontiers in Neuroengineering*, *7*. https://doi.org/10.3389/fneng.2014.00024

Fisher, R. S., Harding, G., Erba, G., Barkley, G. L., & Wilkins, A. (2005). Photic- and Pattern-induced Seizures: A Review for the Epilepsy Foundation of America Working Group. *Epilepsia*, *46*(9), 1426–1441. https://doi.org/10.1111/j.1528-1167.2005.31405.x

Floriano, A., Diez, P. F., & Freire Bastos-Filho, T. (2018). Evaluating the Influence of Chromatic and Luminance Stimuli on SSVEPs from Behind-the-Ears and Occipital Areas. *Sensors (Basel, Switzerland)*, *18*(2). https://doi.org/10.3390/s18020615

Guo, F., Hong, B., Gao, X., & Gao, S. (2008). A brain–computer interface using motion-onset visual evoked potential. *Journal of Neural Engineering*, *5*(4), 477–485. https://doi.org/10.1088/1741-2560/5/4/011

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, *66*(1), 51–83. https://doi.org/10.1109/PROC.1978.10837

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.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Hsu, H.-T., Lee, I.-H., Tsai, H.-T., Chang, H.-C., Shyu, K.-K., Hsu, C.-C., Chang, H.-H., Yeh, T.-K., Chang, C.-Y., & Lee, P.-L. (2016). Evaluate the Feasibility of Using Frontal SSVEP to Implement an SSVEP-Based BCI in Young, Elderly and ALS Groups. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *24*(5), 603–615. https://doi.org/10.1109/TNSRE.2015.2496184

Huang, Q., Zhang, Z., Yu, T., He, S., & Li, Y. (2019). An EEG-/EOG-Based Hybrid Brain-Computer Interface: Application on Controlling an Integrated Wheelchair Robotic Arm System. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.01243

Huggins, J. E., Wren, P. A., & Kirsten L. Gruis. (2011). What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis*, *12*(5), 318–324. https://doi.org/10.3109/17482968.2011.572978

Hwang, H.-J., Kim, S., Choi, S., & Im, C.-H. (2013). EEG-Based Brain-Computer Interfaces: A Thorough Literature Survey. *International Journal of Human–Computer Interaction*, *29*(12), 814–826. https://doi.org/10.1080/10447318.2013.780869

Installation—Anaconda and Miniconda. (n.d.). Psychopy3. Retrieved February 13, 2021, from https://www.psychopy.org/download.html#anaconda-and-miniconda

Jiang, J., Fares, A., & Zhong, S.-H. (2019). A Context-Supported Deep Learning Framework for Multimodal Brain Imaging Classification. *IEEE Transactions on Human-Machine Systems*, *49*(6), 611–622. https://doi.org/10.1109/THMS.2019.2904615

Keller, J., Böhm, S., Aho-Özhan, H. E. A., Loose, M., Gorges, M., Kassubek, J., Uttner, I., Abrahams, S., Ludolph, A. C., & Lulé, D. (2018). Functional reorganization during cognitive function tasks in patients with amyotrophic lateral sclerosis. *Brain Imaging and Behavior*, *12*(3), 771–784. https://doi.org/10.1007/s11682-017-9738-3

Kim, Y. J., Nam, H. S., Lee, W. H., Seo, H. G., Leigh, J.-H., Oh, B.-M., Bang, M. S., & Kim, S. (2019). Vision-aided brain–machine interface training system for robotic arm control and clinical application on two patients with cervical spinal cord injury. *BioMedical Engineering OnLine*, *18*(1), 14. https://doi.org/10.1186/s12938-019-0633-6

Kinney-Lang, E. W. (2018). *Application of multi-way analysis for characterizing paediatric electroencephalogram (EEG) recordings* [Doctor of Philosophy, The University of Edinburgh]. https://era.ed.ac.uk/handle/1842/35591

Kinney-Lang, E. W., Auyeung, B., & Escudero, J. (2016). Expanding the (kaleido)scope: Exploring current literature trends for translating electroencephalography (EEG) based brain–computer interfaces for motor rehabilitation in children. *Journal of Neural Engineering*, *13*(6), 061002. https://doi.org/10.1088/1741-2560/13/6/061002

Kinney-Lang, E. W., Ebied, A., Auyeung, B., Chin, R. F. M., & Escudero, J. (2019). Introducing the Joint EEG-Development Inference (JEDI) Model: A Multi-Way, Data Fusion Approach for Estimating Paediatric Developmental Scores via EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(3), 348–357. https://doi.org/10.1109/TNSRE.2019.2891827

Kübler, A., Holz, E. M., Riccio, A., Zickler, C., Kaufmann, T., Kleih, S. C., Staiger-Sälzer, P., Desideri, L., Hoogerwerf, E.-J., & Mattia, D. (2014). The User-Centered Design as Novel Perspective for Evaluating the Usability of BCI-Controlled Applications. *PLOS ONE*, *9*(12), e112392. https://doi.org/10.1371/journal.pone.0112392

Lazarou, I., Nikolopoulos, S., Petrantonakis, P. C., Kompatsiaris, I., & Tsolaki, M. (2018). EEG-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: A novel approach of the 21st century. *Frontiers in Human Neuroscience*, *12*, 14.

Lotze, M., & Halsband, U. (2006). Motor imagery. *Journal of Physiology Paris*, *86*(4–6), 386–395. https://doi.org/10.1016/j.jphysparis.2006.03.012

Luck, S. J. (2014). An Introduction to the Event-Related Potential Technique (2nd ed.). MIT Press.

Ma, T., Li, H., Deng, L., Yang, H., Lv, X., Li, P., Li, F., Zhang, R., Liu, T., Yao, D., & Xu, P. (2017). The hybrid BCI system for movement control by combining motor imagery and moving onset visual evoked potential. *Journal of Neural Engineering*, *14*(2), 026015. https://doi.org/10.1088/1741-2552/aa5d5f

Mahon, S., Faulkner, J., Barker-Collo, S., Krishnamurthi, R., Jones, K., & Feigin, V. (2020). Slowed Information Processing Speed at Four Years Poststroke: Evidence and Predictors from a Population-Based Follow-up Study. *Journal of Stroke and Cerebrovascular Diseases*, *29*(2), 104513. https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.104513

Matsuura, M., Okubo, Y., Toru, M., Kojima, T., He, Y., Shen, Y., & Kyoon Lee, C. (1993). A cross-national EEG study of children with emotional and behavioral problems: A WHO collaborative study in the Western Pacific region. *Biological Psychiatry*, *34*(1), 59–65. https://doi.org/10.1016/0006-3223(93)90257-E

Meng, J., Zhang, S., Bekyo, A., Olsoe, J., Baxter, B., & He, B. (2016). Noninvasive Electroencephalogram Based Control of a Robotic Arm for Reach and Grasp Tasks. *Scientific Reports*, *6*, 38565. https://doi.org/10.1038/srep38565

Mikołajewska, E., & Mikołajewski, D. (2014). The prospects of brain—Computer interface applications in children. *Open Medicine*, *9*(1), 74–79.

Mussi, M. G. (2021). Development of A Simultaneous Hybrid Brain-Computer Interface Using SSVEP and P300. *RESNA Annual Conference - 2021*.

https://www.resna.org/sites/default/files/conference/2021/NewEmergingTechnology/75_Mussi/75_Mussi.pdf

Nakanishi, M., Wang, Y., Wang, Y.-T., & Jung, T.-P. (2015). A Comparison Study of Canonical Correlation Analysis Based Methods for Detecting Steady-State Visual Evoked Potentials. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0140703

Okudan, Z. V., & Özkara, Ç. (2018). Reflex epilepsy: Triggers and management strategies. *Neuropsychiatric Disease and Treatment*, *14*, 327. https://doi.org/10.2147/NDT.S107669

Pan, J., Xie, Q., He, Y., Wang, F., Di, H., Laureys, S., Yu, R., & Li, Y. (2014). Detecting awareness in patients with disorders of consciousness using a hybrid brain–computer interface. *Journal of Neural Engineering*, *11*(5), 056007. https://doi.org/10.1088/1741-2560/11/5/056007

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pierce, J. R. (1980). *An Introduction to Information Theory: Symbols, Signals & Noise*. Dover Publications. https://books.google.ca/books?id=fXxde44_0zsC

Punsawad, Y., & Wongsawat, Y. (2017). Enhancement of steady-state visual evoked potential-based brain–computer interface systems via a steady-state motion visual stimulus modality. *IEEJ Transactions on Electrical and Electronic Engineering*, *12*, S89–S94. Scopus. https://doi.org/10.1002/tee.22422

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaManita, A.-S., McNamara, J. O., & Williams, S. M. (2004). *Neuroscience* (3rd ed.). Sinauer Associates.

Report of the committee on methods of clinical examination in electroencephalography (pp. 370–375). (1958). [Technical Report]. Elsevier. https://doi.org/10.1016/0013-4694(58)90053-1

Reyes, J. M. A., & Forgach, C. E. S. (2016). Evaluation of the Minimum Size of a Window for Harmonics Signals. *Journal of Signal and Information Processing*, *07*(04), 175. https://doi.org/10.4236/jsip.2016.74017

Rong, Y., Wu, X., & Zhang, Y. (2020). Classification of motor imagery electroencephalography signals using continuous small convolutional neural network. *International Journal of Imaging Systems and Technology*. Scopus. https://doi.org/10.1002/ima.22405

Sabbah, P., Schonen, S. de, Leveque, C., Gay, S., Pfefer, F., Nioche, C., Sarrazin, J.-L., Barouti, H., Tadie, M., & Cordoliani, Y.-S. (2002). Sensorimotor Cortical Activity in Patients with Complete Spinal Cord Injury: A Functional Magnetic Resonance Imaging Study. *Journal of Neurotrauma*, *19*(1), 53–60. https://doi.org/10.1089/089771502753460231

Saravanakumar, D., & Reddy M., R. (2018). A Visual Keyboard System Using Hybrid Dual Frequency SSVEP Based Brain Computer Interface with VOG Integration. *2018 International Conference on Cyberworlds (CW)*, 258–263. https://doi.org/10.1109/CW.2018.00053

Saravanakumar, D., & Reddy, R. M. (2018). A Novel Visual Keyboard System for Disabled People/Individuals using Hybrid SSVEP Based Brain Computer Interface. In A. Sourin, O. Sourina, C. Rosenberger, & M. Erdt (Eds.), *2018 International Conference on Cyberworlds (cw)* (pp. 264–269). https://doi.org/10.1109/CW.2018.00054

Schudlo, L. C., & Chau, T. (2018). Development and testing an online near-infrared spectroscopy brain–computer interface tailored to an individual with severe congenital motor impairments. *Disability and Rehabilitation: Assistive Technology*, *13*(6), 581–591. https://doi.org/10.1080/17483107.2017.1357212

Singh, V., Rana, R. K., & Singhal, R. (2013). Analysis of repeated measurement data in the clinical trials. *Journal of Ayurveda and Integrative Medicine*, *4*(2), 77–81. https://doi.org/10.4103/0975-9476.113872

Speier, W., Deshpande, A., & Pouratian, N. (2015). A Method for Optimizing EEG Electrode Number and Configuration for Signal Acquisition in P300 Speller Systems. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *126*(6), 1171–1177. https://doi.org/10.1016/j.clinph.2014.09.021

Stadskleiv, K. (2020). Cognitive functioning in children with cerebral palsy. *Developmental Medicine & Child Neurology*, *62*(3), 283–289. https://doi.org/10.1111/dmcn.14463

Supported Boards—BrainFlow documentation. (n.d.). Retrieved February 13, 2021, from https://brainflow.readthedocs.io/en/stable/SupportedBoards.html

Thurlings, M. E., Brouwer, A.-M., Van Erp, J. B. F., & Werkhoven, P. (2014). Gaze-independent ERP-BCIs: Augmenting performance through location-congruent bimodal stimuli. *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00143

Tou, J. T., & Gonzalez, R. C. (1977). *Pattern recognition principles*. Addison-Wesley Pub. Co.; /z-wcorg/. http://books.google.com/books?id=Bb9QAAAAYAAJ

Viera, A. J., & Garrett, J. M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*, 4.

Walter, W. G., Cooper, R., Aldridge, V., McCallum, W., & Winter, A. (1964). Contingent negative variation: An electric sign of sensori-motor association and expectancy in the human brain. *Nature*, *203*(4943), 380–384.

Wolpaw, J. R., & Wolpaw, E. W. (2012). *Brain–Computer Interfaces: Principles and Practice*. Oxford University Press. dali

Wolpaw, J., & Wolpaw, E. W. (2012). *Brain–Computer Interfaces: Principles and Practice*. Oxford University Press.

Yin, E., Zeyl, T., Saab, R., Hu, D., Zhou, Z., & Chau, T. (2015a). A Hybrid Brain–Computer Interface Based on the Fusion of P300 and SSVEP Scores. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *23*(4), 693–701. https://doi.org/10.1109/TNSRE.2015.2403270

Yin, E., Zeyl, T., Saab, R., Hu, D., Zhou, Z., & Chau, T. (2015b). An Auditory-Tactile Visual Saccade-Independent P300 Brain–Computer Interface. *International Journal of Neural Systems*, *26*(01), 1650001. https://doi.org/10.1142/S0129065716500015

Yousefi, R., Rezazadeh Sereshkeh, A., & Chau, T. (2019). Online detection of error-related potentials in multi-class cognitive task-based BCIs. *Brain-Computer Interfaces*, *6*(1–2), 1–12. https://doi.org/10.1080/2326263X.2019.1614770

Zhang, J. Z., Jadavji, Z., Zewdie, E., & Kirton, A. (2019). Evaluating if children can use simple brain computer interfaces. *Frontiers in Human Neuroscience*, *13*, 24.

Zhang, Z., Huang, Y., Chen, S., Qu, J., Pan, X., Yu, T., & Li, Y. (2017). An Intention-Driven Semi-autonomous Intelligent Robotic System for Drinking. *Frontiers in Neurorobotics*, *11*. https://doi.org/10.3389/fnbot.2017.00048

Zina Li, Shuqing Zhang, & Jiahui Pan. (2019). Advances in Hybrid Brain-Computer Interfaces: Principles, Design, and Applications. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2019/3807670

Zuo, C., Jin, J., Yin, E., Saab, R., Miao, Y., Wang, X., Hu, D., & Cichocki, A. (2019). Novel hybrid brain–computer interface system based on motor imagery and P300. *Cognitive Neurodynamics*. https://doi.org/10.1007/s11571-019-09560-x

Appendix 1

RESNA Paper

Mussi, M. G. (2021). Development of A Simultaneous Hybrid Brain-Computer Interface Using SSVEP and P300. *RESNA Annual Conference - 2021*.
https://www.resna.org/sites/default/files/conference/2021/NewEmergingTechnology/75_Mussi/75_Mussi.pdf

Development of a simultaneous Hybrid Brain-Computer Interface using SSVEP and P300

Matheus G Mussi

University of Alberta (Edmonton)

INTRODUCTION

People experiencing neurological disorder can have severe limitations in their functional abilities [1]. They usually rely on assistive technology to perform daily activities [2], but sometimes they are difficult to control. Brain-computer interfaces (BCI) are a potential assistive technology solution. This technology is feasible for people with complex physical needs because in many cases, the disorders do not affect the cognition extensively [3–7]. Traditional BCI methods, nevertheless, yield insufficient performance to be used in real-time applications and are hard to operate independently, which are important criteria for end-users [8,9]. Individuals experience fatigue when low accuracy rates require them to repeat entries and correct selections. Traditional BCI rely on a single input signal (e.g. EEG), single source of stimulus (e.g. auditory, visual, tactile, etc) or a single brain signal paradigm (pattern). This causes the system to have low information transfer rate, an inflexible human-interface and less information to improve its performance [10]. Hybrid brain-computer interfaces aim to improve on performance and speed through multi-modal signal inputs, combining different brain-signals, BCI paradigms or other external device input [11].

There are many hybrid brain-computer interface (hBCI) systems that use various combinations of brain signals and physiological signals or other devices. Nonetheless, very few papers explain the details of how to 'assemble' an hBCI. This paper explains the development of software capable of displaying simultaneous stimuli of SSVEP and P300 which will be used in future research with children with disabilities.

ARCHITECTURE

Interface

For our system's display, there will be three squares that have coloured centre areas that flash at different frequencies for the SSVEP stimuli, and an outline edge that will appear around the squares one at a time in a pseudo-random order for the P300 stimulus, as shown in Fig. 1.

For the SSVEP, squares will create the flashing effect by interpolating between colors, as proposed by [12]. The chosen frequencies are 30, 10 and 6 Hz, to avoid seizure-inducing ranges. When the classification is concluded, the selected square's centre area will briefly turn white in colour to indicate the classifier chose that square.

Program

For the online operation of the system, threads are needed to avoid delays in the execution of processes that need to be concurrent. The architecture of the program is based on a main section that initializes other threads and variables, the experiment thread, which displays the squares on the screen using Psychopy, the acquisition thread, the feature extraction threads (SSVEP and P300 specific) and the classification threads. The overall architecture scheme can be seen in Fig. 2.

To synchronize the functioning of the different paradigms, the program is based on threading event and standardized intervals. Every half second, a new P300 stimulus is presented and the buffered data is pre-processed and sent to the classifier for training or validation. Events are sent to the active threads and their function is synchronized, avoiding data loss. An important

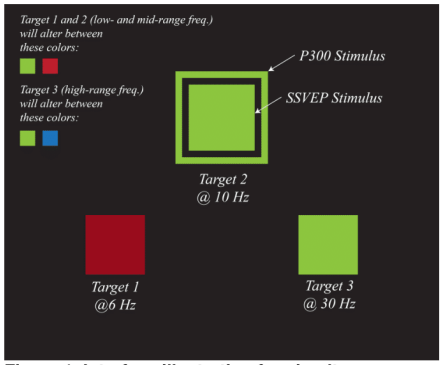

Figure 1. Interface illustration for simultaneous SSVEP and P300 paradigms

note on threading events is that they have a built-in *wait* function, which avoids unnecessary loops and protects Python's synchronization primitives.

To transfer data from one thread to the other, threading queues are used. They are needed because Global variables are not supported by threads. Queues have a first-in-first-out method and they are emptied every time the *get* method is used.

DEVELOPMENT

The developed system is based on Python3®. The libraries used are freely available online and the software should be compatible with many BCI headsets, depending primarily on the BrainFlow library. This system was tested with the OpenBCI Cython board. The system runs in a Windows 10 operational system, but most of the installations should also work with Linux. Mac users might experience some compatibility issues with some headsets (especially due to the deprecation of Future Technology Devices International (FTDI) chip drivers, which hinders the use of the OpenBCI board).

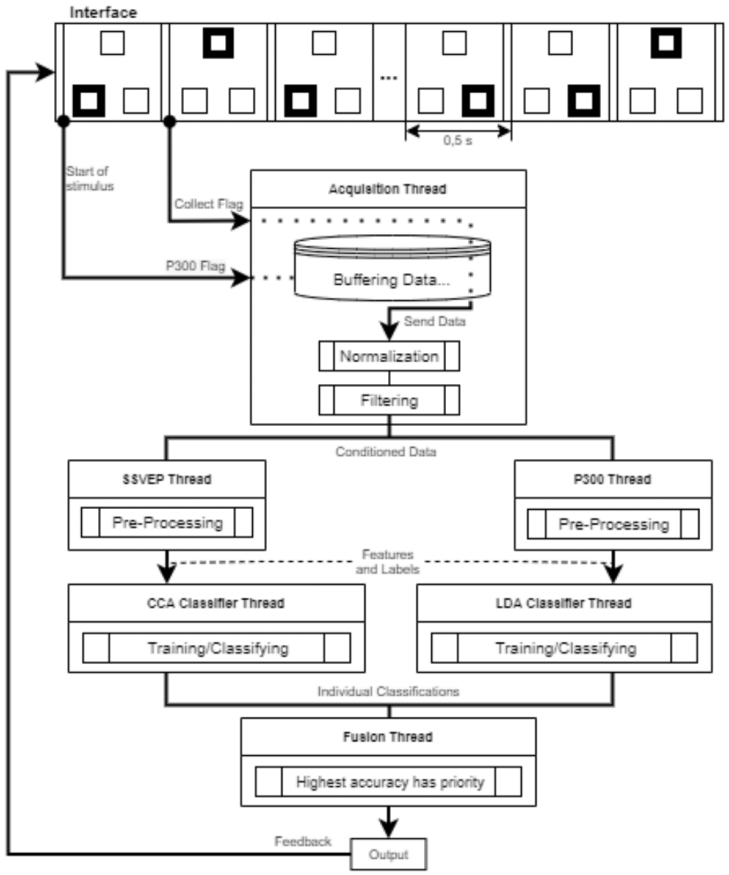

Figure 2. Overview of the program architecture scheme

Setting up the experiment on a new computer

This sub-section explains the installation of all the necessary resources to create a hBCI system with visual stimulation. The steps below are meant to be done in sequence.

Requirements

The basic requirements for the system are: 1) Python 3.6.6, 2) Anaconda3 and 3) Psychopy3. Python 3.6.6 is currently the most up-to-date version that supports all the required packages. In addition to Python, Anaconda was used to maintain all the specific libraries for this system separated from other libraries in the operational system. Psychopy3 is a package used to manipulate screen elements to create stimuli for the users.

Creating an environment with Anaconda with Psychopy3

Environments are directories containing specific packages which do not affect other environments. This gives the ability of installing packages and package versions that are project-specific [13]. For this project, the environment is based on an environment containing Pychopy3. This method avoids compatibility issues.

Firstly, the Anaconda3 executable [14] and the Psychopy3 environment file (psychopy-env.yml) [15] need to be downloaded. Anaconda needs then to be installed and the Psychopy file saved in an accessible folder. Once

anaconda3 is installed, through the Anaconda Powershell, navigate to the folder containing the Psychopy file and enter `conda env create -n psychopy -f psychopy-env.yml` to create the environment named "psychopy". To activate the environment, in the Powershell, enter `conda activate psychopy`.

Once the environment is running, some packages are needed (some packages are OpenBCI Cython-specific). Most of them need to be installed via `conda install -c` command, but some are installed through `pip install`. The packages are `cython`, `pyserial`, `matplotlib`, `piglet` (v.1.4.10), `numpy`, `pygtgraph`, `scipy`, `pandas`, `pylsl`, and `Pillow`.

The specific Pyglet version needs to be 1.4.10 instead of newer versions because some later versions yield the error "WMF Failed to initialize threading: err.strerror" when using threads. Depending on the application, some dll files can be missing and also generate errors. If the errors "The program can't start because VCRUNTIME140D.dll is missing from your computer" or "The program can't start because ucrtbased.dll is missing from your computer" appear, the dll's need to be downloaded and pasted in the C:/Windows/System32 folder.

Installing BrainFlow

BrainFlow is a library to parse and obtain biosensory data from devices. This library allows an easy connectivity between electroencephalography (EEG) headsets and the computer/program. The strengths of this library over others are its easy implementation, its board versatility (e.g. a few lines of code can easily adapt the program to receive signals from OpenBCI Cython or g.tec Unicorn) and its vivid community, which gives vast support through Slack. To install BrainFlow in the environment, enter `python -m pip install brainflow`.

Installing Scikit-learn

Scikit-learn is a library for predictive and data analysis. It provides simplicity and a vast range of tools to classify and analyse data. The classifiers in this library will be implemented for the system's decision making. It can also be installed in the environment with `conda install scikit-learn`.

Remove excessive signal latency

In some boards, the default latency can cause the signal to be read incorrectly or in chunks. To avoid these, the latency needs to be decreased. This can be done by going to the *Device Manager* (Right click on *My Computer/This PC*, then *Properties* and select *Device Manager*). Then, the port setting needs to be adjusted: expand the *Ports (COM & LPT)* list, right click the serial port corresponding to the device (e.g. COM3) and select *Properties*. In *Properties*, go to the *Port Settings* tab, then *Advanced,* and then change *Latency Timer (msec)* to 1.

Working with BrainFlow

BrainFlow gives the developer the ability to connect and start streaming from within the code. The minimum sequence to connect to a board is the following:

```
import brainflow
from brainflow.board_shim import BardShim, BrainFlowInputParams, BoardIds
from brainflow.data_filter import DataFilter, FilterTypes, AggOperation
BoardShim.enable_dev_board_logger()
params = BrainFlowInputParams()
params.serialport = 'COM3'
board_id = BoardIds.CYTON_BOARD.value      #variable for specific board
board = BoardShim(board_id, params)        #creates callable object
board.prepare_session()                    #finds and connects to board
board.start_stream()                       #starts streaming
data = board.get_board_data()              #collects data
```

Each board needs a specific set of parameters. To see what parameters are needed, refer to [16]. For the `board_id` variable, the board in use should be called (e.g. if using the g.tec Unicorn, use `board_id = BoardIds.UNICORN_BOARD.value`). The data streamed from the board comes in a single array which contains timestamps and sensory data.

FUTURE RESEARCH

This hBCI is currently undergoing testing and the next steps are 1) tests with adults with and without disabilities, using a User Centred Design approach to improve the system and 2) experiments with children with disabilities to see its efficacy in clinical applications.

REFERENCES

1. Bauer G, Gerstenbrand F, Rumpl E. Varieties of the locked-in syndrome. J Neurol. 1979 Aug 1;221(2):77–91.

2. Cook AM, Polgar JM. Assistive technologies : principles and practice [Internet]. 2014. Available from: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1151635

3. Stadskleiv K. Cognitive functioning in children with cerebral palsy. Dev Med Child Neurol. 2020;62(3):283–9.

4. Mahon S, Faulkner J, Barker-Collo S, Krishnamurthi R, Jones K, Feigin V. Slowed Information Processing Speed at Four Years Poststroke: Evidence and Predictors from a Population-Based Follow-up Study. J Stroke Cerebrovasc Dis. 2020 Feb 1;29(2):104513.

5. Cioni G, Guzzetta A, D'Acunto G. Cerebral Plasticity and Functional Reorganization in Children with Congenital Brain Lesions. In: Buonocore G, Bracci R, Weindling M, editors. Neonatology: A Practical Approach to Neonatal Diseases [Internet]. Milano: Springer Milan; 2012 [cited 2020 Sep 9]. p. 145–9. Available from: https://doi.org/10.1007/978-88-470-1405-3_21

6. Keller J, Böhm S, Aho-Özhan HEA, Loose M, Gorges M, Kassubek J, et al. Functional reorganization during cognitive function tasks in patients with amyotrophic lateral sclerosis. Brain Imaging Behav. 2018 Jun;12(3):771–84.

7. Sabbah P, Schonen S de, Leveque C, Gay S, Pfefer F, Nioche C, et al. Sensorimotor Cortical Activity in Patients with Complete Spinal Cord Injury: A Functional Magnetic Resonance Imaging Study. J Neurotrauma. 2002 Jan;19(1):53–60.

8. Huggins JE, Wren PA, Kirsten L. Gruis. What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis. Amyotroph Lateral Scler. 2011 Sep 1;12(5):318–24.

9. Collinger JL, Boninger ML, Bruns TM, Curley K, Wang W, Weber DJ. Functional priorities, assistive technology, and brain-computer interfaces after spinal cord injury. J Rehabil Res Dev. 2013;50(2):145.

10. Zina Li, Shuqing Zhang, Jiahui Pan. Advances in Hybrid Brain-Computer Interfaces: Principles, Design, and Applications. Comput Intell Neurosci [Internet]. 2019 [cited 2020 Feb 23]; Available from: https://www.hindawi.com/journals/cin/2019/3807670/

11. Wolpaw J, Wolpaw EW. Brain–Computer Interfaces: Principles and Practice. Oxford University Press; 2012. 424 p.

12. Floriano A, Diez PF, Freire Bastos-Filho T. Evaluating the Influence of Chromatic and Luminance Stimuli on SSVEPs from Behind-the-Ears and Occipital Areas. Sensors [Internet]. 2018 Feb 17 [cited 2020 Oct 8];18(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5855130/

13. Conda environments [Internet]. [cited 2021 Feb 13]. Available from: https://docs.conda.io/projects/conda/en/latest/user-guide/concepts/environments.html

14. Anaconda - Individual Edition [Internet]. Anaconda. [cited 2021 Feb 13]. Available from: https://www.anaconda.com/products/individual

15. Installation - Anaconda and Miniconda [Internet]. Psychopy3. [cited 2021 Feb 13]. Available from: https://www.psychopy.org/download.html#anaconda-and-miniconda

16. Supported Boards - BrainFlow documentation [Internet]. [cited 2021 Feb 13]. Available from: https://brainflow.readthedocs.io/en/stable/SupportedBoards.html

Appendix 2

Studies Table

Study (N = 17)	Age	Pop. Size	Condition	Paradigm	Classifier	Activity	Accuracy [%]
(Meng et al., 2016)	18 to 54	13	wo/ disability	MI	ERS/ERD filter	Robotic arm	95
(Cincotti et al., 2008)	12 to 35	4	w/ disability	MI	Statistical Analysis	Cursor movement	66.57
(Huang et al., 2019)	22 to 37	5	wo/ disability	MI	SVM	Wheelchair	88
(Pan et al., 2014)	16 to 70	8	w/ disability	P300 + SSVEP	SVM + Power Ratio Detection	Option selection	60.52
(Yin et al., 2015a)	18 to 35	13	wo/ disability	P300 + SSVEP	SWLDA + CCA	Speller control	95.18
(Yin et al., 2015b)	20 to 28	12	wo/ disability	P300*	BLDA	Option selection	88.67
(Zuo et al., 2019)	22 to 28	18	wo/ disability	MI + P300	BLDA	Option selection	93.94
(Carmona et al., 2020)	29 ± 5	15	-	SSEP*	CCA	Option selection	85
(Thurlings et al., 2014)	22 to 26	10	wo/ disability	P300*	SWLDA	Option selection	85
(Saravanakumar & Reddy M., 2018)	21 to 31	10	wo/ disability	SSVEP	Extended Multivariate Synchronization index	Speller control	94.99
(Z. Zhang et al., 2017)	19 to 21	8	wo/ disability	P300	BLDA	Robotic arm	97.5
(Choi et al., 2018)	21 to 24	5	wo/ disability	SSVEP	CCA	Chess game	85.8

(J. Z. Zhang et al., 2019)	6 to 18	26	wo/ disability	MI	PNN and Radial Basis Function	Car movement and Cursor movement	Kappa: 0.46
(Kim et al., 2019)	37 and 47	2	w/ disability	MI	Artificial Potentials	Robotic arm	57.37
(Ehlers et al., 2012)	G1: 6,73 G2: 8,08 G3: 9.86 G4: 22,36	11 12 14 14	wo/ disability	SSVEP	Bremen Algorithm	Speller control	58 53 75 78
(Cho et al., 2017)	24.8 ± 3.86	52	wo/ disability	MI	LDA	Option selection	67.46
(Yousefi et al., 2019)	29 ± 3	10	wo/ disability	non-MI	rLDA	Option selection	67

*Note. Paradigms with * mean they were multisensory stimuli. Ehlers et al. (2012) presented results for four different age groups.*

Appendix 3

Session Protocols

Protocols	Steps	Amount of time [min]	Check
Cleaning – pre-session	Sanitize hands (with Isopropyl alcohol or water and soap)	0.5	□
	Sanitize chairs and desks where participant and family members are going to sit	1	□
	Cap and electrodes will be previously sanitized	0	□
Welcome	Health assessment (following the AHS protocols): a) If it is the first session, collect contact information (from participants or family member(s)) b) Register date of experiment for tracking purposes c) If assessment indicates infection for participant or family member(s), session will be terminated	5	□
	If it is the first session, explain procedures and experiment objectives	5	□
Technical	Turn EEG device on and verify signal acquisition is working properly: a) If not, reboot device b) If problem persists, reboot computer	2	□
	Put electrodes (cap) on participant's head	10	□
	Adjust cap a) Pull and push cap until it reaches the right electrode spots b) Apply more gel where needed c) Tighten or loosen the cap to make the participant comfortable	2	□
	Verify impedance of each electrode. Repeat items above if: a) any electrode is below 5 ohms or above 100 ohms b) participant feels uncomfortable c) electrodes are out of place in relation to the 10/20 placement system	1	□
Session	Run a practice run where no results will be saved:	2 to 5	□

	a) Check after each practice run if participant feels comfortable with the system. Repeat as needed.		
	Run the classifier training set: a) Set up video camera and hit record b) Turn on data collection c) Check if raw EEG data and results were stored in hard drive after first run d) After each run, allow a minute break if needed e) After the training is successfully completed, run the classifier training script	5 to 15	□
	Run the four feedback sets: a) After each run, allow a minute break if needed b) After each set, allow a 5 minutes break if needed	20 to 60	□
	Remove cap from participant's head	0.5	□
	Store data: a) Label data with date and participant code b) Add description stating any unexpected events or behaviours, if any	2	□
	Ask for participant member(s) opinion: a) fill the NASA TLX survey b) annotate any other verbal opinion/suggestion for the system or experiment	5	□
Cleaning – post-session	Clean cap with water and detergent	5	□
	Brush electrodes to remove gel residues with water and detergent	2	□
	Sanitize used chairs and desks	1	□
Preparation Time		11.5	
Participant Time		57.5 to 110.5	
Total		69 to 122	

Appendix 4

Procedures and Approvals due to COVID-19

Some extra procedures need to be considered for the safety of the researchers and the participants due to COVID-19. The following list of requirements were followed in accordance with the restrictions and guidance required by the University of Alberta:

- Before each session, when entering the facility, a screening process will be done to ensure participants and parents or companions are feeling well
- Contact information will be collected and stored for the duration of the study plus two weeks afterwards, in case an infectious case needs to be traced
- Researchers and companions will wash their hands before and after contact with participants
- Instructions on how to cough and sneeze safely will be reinforced
- Personal protective equipment, namely masks, will be distributed for those who do not have one
- At all times possible, social distancing will be maintained (2m apart). In situations where contact is needed, both participant and researcher will be using masks, and researchers will use protection googles.

  Before and after sessions, the chairs and equipment that were used by the participant and companions will be cleaned and disinfected with isopropyl alcohol.

Approvals

Special approvals and documentation were required to return to research activities during the COVID-19 pandemic. The following approvals were granted before commencing the experiments:

- University of Alberta
    - Ethics approval from the Research Ethics Office to suit safety measures
    - Environment Health and Safety (EHS) Return to Campus Plan

Appendix 5

Notes taking during sessions

The following table displays the notes and incidental comments registered during sessions. The comments were made by the participants about the system during the sessions. The comments were reproduced *verbatim*, except where indicated.

	Sess.	Paradigm	Notes	Incidental Comments
P1	1	P300	Feeling sleepy after lunch. Struggling to stay awake.	-
P1	4	Hybrid	-	*It's a little harder, but not too bad* (referring to the hybrid compared to the others) I'm sorry, I think I counted one of them [P300 frames] wrong
P2	1	SSVEP	Slept only 5 hours previous night. Struggling to stay awake.	It is a lot of flashing, not going to lie*
P2	3	Hybrid	-	The flashing part puts me to sleep a little bit*
P2	4	SSVEP	-	This time it was nice!
P2	6	P300	Tired Struggling to stay awake	-
P3	1	SSVEP	Reference fell off between sets 2 and 3, and was reattached.	-
P3	4	Hybrid	-	Should it [the counting] always be 7? I think I got 6.
P3	5	P300	Tired. Struggling to stay awake.	-
P3	6	SSVEP	Slept only 4 hours previous night. Struggling to stay awake from sets 1-3.	-
P4	3	Hybrid	-	It is not that bad, but very similar to the first one [SSVEP]
P4	4	Hybrid	Struggling to stay awake during set 3. Feeling really tired during sets 3-5.	I counted six this time
P4	5	SSVEP	-	I feel this square [10 Hz] is the easier one to focus [on]
P5	1	SSVEP	Noise in the hallway during set 2. Used earbuds to block sound from sets 3-5.	-
P5	3	Hybrid	Longer break between sets 3 and 4 due to tiredness.	There is way more going on.

				I think I will take a longer break this time (after being asked if he needed a break)
P5	5	Hybrid	Feeling more tired than usual, especially from sets 3-5.	I definitely got one of them wrong [miscounted]*
P5	6	SSVEP	Feeling more tired than usual. Struggling to stay awake during set 3.	-
P6	3	Hybrid	-	I can still count well, but it's more stuff at the same time

*Note. *: adapted comments.*