Using an Electroencephalography Brain-Computer Interface for Monitoring Mental Workload During Flight Simulation

by

Adam Fraser

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Cognitive Science

In

Cognitive Science

Carleton University

Ottawa, Ontario

Abstract

Background: The objective of the present research was to investigate the potential of implementing electroencephalography (EEG) in a passive brain-computer interface for monitoring mental workload during virtual reality flight simulation. Most aviation accidents are related to pilot cognition and a mismatch between task demands and cognitive resources. Real-time neurophysiological monitoring that identifies high-workload mental states offers an effective approach for reducing accidents during flight.

Method: Non-pilot participants performed simulated flight operations. Workload was manipulated to represent regular flight scenarios by varying navigational difficulty and performing communication tasks. EEG data was collected and used to classify periods of flight as high or medium workload.

Results and implications: A classification rate of 75.9% was obtained which provides promise for future use of EEG brain-computer interfaces in aviation practice. The most informative classification features (Alpha and Beta oscillations) may represent components of working memory which corresponds to predictions from a multiple resource theory approach to experimental design.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Motivation for the Present Work

General aviation has a fatal accident rate more than twenty times greater than that of scheduled airline operations (Bureau of Transportation Statistics, 2019; Federal Aviation Administration, 2018). The high accident rate in general aviation has been attributed, in part, to events where task demands exceed a pilot's cognitive capacity (AOPA Foundation, 2016; Wickens, 2002). The use of neurophysiological monitoring to objectively classify a pilot's cognitive state during flight may provide a way to monitor workload and provide timely feedback to pilots. The objective of the present research was to investigate the potential of implementing electroencephalography (EEG) in a passive brain-computer interface (BCI) for monitoring mental workload during virtual reality (VR) flight simulation. Mental workload was manipulated following a multiple resource theory of workload (Wickens, 1984; 2002) and manipulations were designed to be representative of relevant variations in workload during regular flight.

1.2. Mental Workload

Mental workload is framed around the assumption that an operator has a limited supply of attentional resources (Wickens, 2002). Performance on one or more tasks will be compromised if the task demands exceed the availability of attentional resources (Kahneman, 1973; Lavie et al., 2004; 2014; Navon & Gopher, 1979). Support for the limited resources framework has both a cognitive and biological basis. At the foundational level, mental resources have been linked to neurobiological constraints. For example, neurons have limited reserves of adenosine triphosphate (ATP) which is a critical energy source for regular neuronal function (Saravini, 1999). Through repetitive activation of neurons, ATP expenditure outweighs ATP

production. ATP is critical for restorative processes such as recovery of resting membrane potential through active ion transportation. The resting membrane potential is essential for electrical signaling in the brain. If the resting membrane potential is not restored due to insufficient ATP stores, further action potentials may be suspended. Although neural processing is metabolically expensive, there is evidence that the brain is supplied with a fixed amount of cellular energy regardless of task load (Bruckmaier et al., 2020; Clarke & Sokoloff, 1999). For example, Bruckmaier et al. demonstrated that levels of a metabolic marker (cytochrome c oxidase) increase in cerebral areas associated with processing attended stimuli and decrease in regions related to processing unattended stimuli. Additionally, marker concentrations increase in task-relevant regions with increased task load. The authors postulate that, through selective attention, cellular energy was being redistributed in accordance with task demands to optimize the use of limited resources. The increase of cellular metabolism and electrical signaling within task-relevant networks can be detected and used to identify relevant brain regions using neuroimaging techniques including electroencephalography (EEG).

1.2.1 Indexing mental workload using EEG brain-computer interfaces.

Electroencephalography is a non-invasive and temporally precise neuroimaging device and a very popular instrument for researching mental states (Abreu et al., 2018). The temporal resolution of EEG (250 Hz+) enables researchers to capture rapid changes in brain activity. EEG has been used to study the neurophysiological correlates of mental workload through inducing, amongst other things, task complexity, working memory load, and attentional load. A summary of EEG research on workload is illustrated in Table 1. Since the 1990's, efforts have been made to integrate EEG into brain-computer interfaces (BCI) for workload level classification (e.g., Pope et al., 1995).

Brain-computer interfaces have traditionally been used to enable users to perform a given task through detection of a physiological correlate of a mental command. BCI systems integrate technology, such as EEG, and data analytics such as machine learning to acquire and interpret the relevant physiological signals and to translate them into the relevant output or action. BCI grew in popularity as achievements were made in domains such as neuroprosthesis, where quadriplegic patients could train control over robotic limbs or computer cursors (e.g., BrainGate, 2005). Other advancements are being made in 'passive' BCI, where the BCI system functions to classify a mental state, without training the individual to produce a particular signal or output (e.g., Dehais et al., 2019; Pope et al., 1995). Passive BCI is the focus of the present research, where the objective is to evaluate the success in detecting high-workload events through EEG activity.

In Pope et al.'s (1995) seminal experiment, participants alternated between automatic (low workload) and manual (high workload) conditions while performing a computerized tracking task from the Multi-Attribute Task Battery (Comstock & Arnegard, 1992). The tracking task was meant to simulate an aviation flight control task. Pope et al. showed that real-time measurements of neuro-oscillatory activity were sensitive to changes between automatic and manual behavior.

Additionally, Pope et al. compared the performance of several proposed EEG indices of workload. Oscillatory patterns of synchronous neural activity are commonly selected features for workload classification, as in most contexts relevant changes in workload are often periodic rather than transient stimulus-driven responses that can be feasibly evaluated using methods such as event-related potentials. Different frequencies are typically categorized into bands of ranges due to associations with mental functions that occur within these bounds. The bands are

commonly defined as Delta 1-4Hz; Theta 4-8Hz; Alpha 8-12Hz; Beta 12-30Hz; & Gamma 30-100Hz. Pope et al. observed that an increase in Beta power and decreases in Alpha and Theta provided a ratio that best approximated workload state. The *Beta / Alpha + Theta* ratio was coined the 'engagement index' and has later been implemented in determining engagement level in action shooter videogames (McMahan et al., 2015) and as an index for mental effort (Rojaz et al., 2020).

The engagement index has been implemented successfully without specification of brain regions. McMahan et al. (2015) calculated values for the engagement index by averaging over all available 14 electrodes, and Pope et al. (1995) obtained index values from a minimal sensor array that provided measurements only from central to posterior regions of the cortex. Although the engagement index specifies an overall decrease in Theta power, increases of Theta power have been observed at specific electrode locations under increased workload. At frontal electrodes increased Theta power has been observed in response to changes in workload including increased task complexity (e.g., steps to solve a puzzle Radüntz, 2020), increased working memory load (Borghini et al., 2014; Gevins et al., 1998; Radüntz, 2017) and during sustained concentration (Paus et al., 1997).

Workload effects on Alpha power are shown to be specific to electrode location. For example, Alpha oscillations are suppressed under increased workload, and are most prominent in electrodes placed over parietal and occipital areas of the cortex (Puma et al., 2018; Wang et al., 2018). Parietal Alpha power has been postulated as an index of workload, particularly in relation to complex problem-solving tasks such as puzzles (Radüntz, 2017; 2020).

Increases in Beta power have also been observed to be relatively strong in parietal and occipital areas during tasks related to visual working memory (Mapelli & Özkurt, 2019).

Increased Beta power has also been linked to visual attention (Wróbel, 2000), short-term memory (Palva et al., 2011; Tallon-Baudry & Bertrand, 1999), and in response to increased working memory demands (Chen & Huang, 2016). Although there has been less mental workload research evaluating Delta frequencies, they have been reported to reflect working-memory load information in frontal regions, and typically reduce in power with increased workload (Zarjam et al., 2011). Due to the anatomical relevance of workload indices, benefits in EEG-BCI could be achieved through evaluating power densities at specified electrodes. A review of existing literature (summarized in Table 1) supports the decision to collect oscillatory EEG measurements for workload monitoring, and to focus on specific frequency ranges and electrode locations.

Table 1

A Summary of EEG Oscillatory Features Associated with Executive Cognitive Functions That Have Been Related to Workload

Beta / Alpha + Theta (Engagement Index)	Sum of all electrodes	Task Engagement: - Pope at al., 1995 - McMahon, 2015
Increased Theta	Frontal electrodes	Task Difficulty: - Gevins et al., 1998 - Radüntz, 2016 Working Memory: - Borghini et al., 2014 Sustained Attention: - Paus et al., 1997
Decreased Alpha	Sum of electrodes, parietal and occipital	Task Engagement: - Pope at al., 1995 - McMahon, 2015 Task Difficulty: - Dasari et al., 2017 - Puma et al., 2018
Increased Beta	Sum of electrodes, parietal, and occipital	Task Engagement: - Pope at al., 1995 - McMahon, 2015 Working Memory: - Mapelli & Özkurt, 2019 - Tallon-Baudry et al., 1999 - Chen and Huang, 2016 - Palva et al., 2011 Visual Attention: - Wróbel, 2000, short-term
Decreased Delta	Frontal electrodes	Working Memory: - Zarjam et al., 2011

Much of the EEG workload literature and examples contained in Table 1 have been conducted in controlled laboratory environments and on computerized tasks. Therefore, it is not clear whether EEG workload signals are robust enough to transfer over into more complicated, non-computerized and non-stationary tasks. Several BCI experiments have been done in-flight or in simulated flight. Dehais et al. (2019) achieved 71% classification accuracy of pilot mental

workload in-flight. Given that the choice of recording instrument was a 6-dry electrode EEG system, signal acquisition was likely far from optimal performance. Wet electrodes and a larger sensor array are known to have substantial contribution in EEG signal quality (Tăuțan et al., 2013). Without the limitations of signal acquisition and sensory array, it is likely that greater classification accuracy can be achieved.

Dehais et al. (2019) classification accuracy was achieved using EEG frequency band features. In contrast, other common neural indices of workload, such as event-related potentials (ERP), were no better at predicting workload than chance. Due to Dehais et al. incorporation of an aircraft, there was likely a degree of electrical interference from turbines and aircraft vibrations that limited the quality of EEG recording. The results suggest that ERP signals cannot withstand the recording limitations of flight environments. A possible explanation for the comparative resilience of oscillatory features is that they aggregate across a time interval, which attenuates the influence of an artefact, whereas temporally specific features such as an ERPs can only attenuate the artefact by averaging across measurements.

A limitation of Dehais et al. (2019) is the compromised ecological validity of the experimental methods that compared a no-load with a high-load condition. In the low-workload condition the evaluated pilot was removed from controlling the aircraft and was instructed to passively observe as a safety co-pilot operated the plane. Removing the load during workload monitoring increases the difference between workload levels relative to regular flight, which would bolster classification potential. Additionally, a no-load / resting condition may relate to qualitatively different neurological states than low or moderate workload, which could further differentiate the classification features. Since smaller differences in workload are expected during real flight, the Dehais et al. findings may not extend into practice. The value of a BCI

system would be in identifying moderate changes in workload, as pilots are always engaged in tasks to some extent.

Flying an aircraft involves a complex environment in comparison to laboratory research. The changing environment of flight simulation induces a variety of responses and mental states. Due to this gap in complexity, it is likely that some phenomena that can be detected in the lab will not be observable in simulated flight. For example, workload indices that relate to specific brain regions, such as Theta and Delta power densities at frontal electrodes, may be compromised by the dynamics of a flight environment. Without distributing the measurement across electrodes, the index may be more susceptible to unexpected phenomena. For example, the effect at a specific electrode could be compromised by changes in signal conductance, displacement, or cognitive events that cannot be controlled for, whereas distributed signals may attenuate the influence of unexpected events by aggregating the measurement across electrodes. Therefore, broader indices such as the engagement index may be expected to perform best in a flight environment.

Moving towards physiological measurement of workload has the benefit over indexing workload through subjective reports, because physiological indexing may be able to overcome the limitations of subjective reporting. The main issue with relying on a subjective report for workload assessment is that subjective reports cannot be feasibly implemented as an aid for evaluation of workload in "real-world" environments. Providing a subjective report is itself another task for the operator, which may be consequential during periods of high workload. Secondly, grounding workload evaluation in subjective reports presumes that an operator has the capacity to accurately assess their level of experienced workload and report it in accordance with a rating scale. Dissociations between subjective ratings of workload and task performance have

raised concerns about the validity of subjective workload questionnaires (see Matthews et al., 2019 for a review). Meanwhile advancements are being made in objective, physiological approaches to monitoring mental workload (Vanneste et al., 2020).

In summary, EEG shows promise as a safety aid for aviation through mental state assessment. Previous research has already revealed associations between EEG features and level of mental workload. However, some EEG workload correlates may have less promise due to anatomical specifications and context dependence.

1.2.2. Workload and multiple resource theory

In his Multiple Resource Theory (MRT), Wickens (1984; 2002) provides a framework for predicting resource sharing between processing systems. Wickens proposed that there exist multiple pools of limited attentional resources that are differentially utilized depending on the sets of tasks being performed. Processing and execution of a task can be delayed or abolished if another concurrently performed task requires the same processing resources. For example, the performance of two concurrent tasks will be degraded less if the tasks depend on different input modalities (Wickens, 2002). When individuals perform two tasks, such as reading instructions and performing a visual search, task processes will have to be performed sequentially as they both demand the visual input modality. Alternatively, reading instructions and listening for an auditory cue involve different inputs and may not need to be performed sequentially.

Wickens et al. (1988) performed simulated helicopter flight missions with secondary cognitive tasks that varied in perceptual processing demand and response modality. The study used several models to predict the impact of the secondary task on navigational performance. The models included different calculations of overall task demand as well as a model that accounted for expected resource sharing. The results indicate that the multiple resources model

was best at predicting performance and accounted for 22% of the variance in vertical deviation from flight path. MRT has also been used in automobile simulator studies (Horrey & Wickens, 2004), where an MRT derived 'interference score' strongly corresponded to the effect a concurrent task would have on driving performance.

Wickens (2002) Multiple Resources Theory is illustrated in Figure 1, where the resource structure is modelled in three dimensions. Figure 1 shows how the processing stage dimension is broken into three categories, each of which contains resources which correspond to cells in the resource structure in the figure. Every resource (or cell) has been dissociated to some degree through dual-task performance measures. At the responding stage (i.e., execution of an action), manual and verbal responses use different resources (Liu & Wickens, 1992; Wickens & Liu, 1988). The resources required at the responding stage use different resources than at the cognitive and perceptual levels. The cognitive and perceptual levels use different resources depending on whether an individual is processing spatial or verbal information (Baddeley, 1986); Wickens refers to spatial and verbal information as processing "codes". The modality dimension is nested within perception and indicates different processing resources for visual and auditory perception of stimuli. Given the multitasking nature of flight, multiple resource models can provide a framework for evaluating flight task difficulty and pilot mental workload.

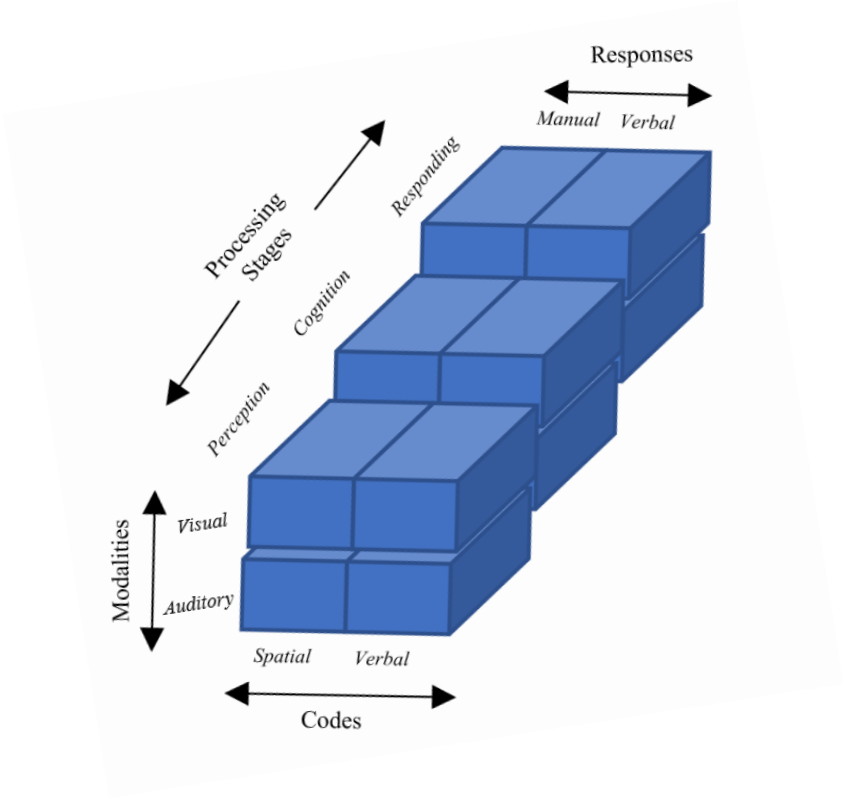

Figure 1. Wickens' (2002) Multiple Resources Theory of workload. The figure displays the three-dimensional structure of mental resources. Note: Wickens (2008) upgraded this model to four dimensions which further separates the visual modality as ambient and focal perception.

1.2.3. Pilot mental workload

During flight, a pilot must engage with a variety of tasks. Tasks include communicating with air traffic control, management of fuel, cabin pressure and electricity, and monitoring of flight instruments, all while performing the primary task of handling the aircraft. As the environment changes, pilot tasks vary in difficulty requiring different degrees of attentional resources to different combinations of tasks. Pilots are trained to prioritize tasks in the order of aviating, navigating, communication and systems management. However, flexibility in priority of tasks is necessary, particularly in the case of unexpected events. To adapt to novelties in flight, the pilot must also engage in decision making and planning to assess the adjusted importance of each task. All of these procedures are to maintain adequate situational awareness

for safe flight, which also encompasses attention to the array of cockpit instruments (Figure 2), routine flight parameters (e.g., attention to the forward path), and to the peripheral environment to safeguard unexpected hazards.

Systems Management: fuel, cabin pressure, electricity etc.
Communications: air traffic control and flight deck.
Systems Awareness: awareness to what 'mode' automated systems are in.
Task Management: understanding what tasks to prioritize at a given moment, and how to navigate between them.

Figure 2. The cockpit of a Cessna 172 general aviation aircraft (left) and outline of some factors of situation awareness in general aviation taken from Wickens (2002). The text box summarizes some of the attentional demands that engage pilots during flight. Photo taken from VirtualPilot3D.com.

Aviation research has shown that performing multiple tasks simultaneously can degrade performance on one or all tasks. For example, increasing workload through demanding attention to multiple surrounding aircrafts impairs a pilot's ability to complete radio calls at specified locations during simulated flight (Van Benthem et al., 2015). Additionally, increasing the amount of visual information to attend to in the flight display decreases pilot's performance on auditory processing tasks (Wanyan et al., 2014). Findings of dual-task performance degradations correspond to the limited resources theory, where increases in visual and auditory perceptual load deplete attentional resources and impairs performance.

Increasing perceptual load has also been shown to interfere with working memory performance (Tsubomi et al., 2013; Van Benthem et al., 2015). Limitations in working memory have been suggested to be related to attentional resources (Cowan et al., 2005). Working memory is related to execution of several flight tasks, and is required to serve among other

things, maintenance of goal-oriented information (Oberauer, 2019), prospective memory (Van Benthem et al., 2015), and planning (Radüntz, 2020). An example of a working memory load during flight would be receiving instructions from air traffic control and maintaining the information in working memory while concurrently maintaining the parameters and strategy necessary to achieve a navigational target.

1.3. Present research

The objective of the present research was to investigate an EEG-BCI for monitoring mental workload during VR flight simulation. Mental workload was manipulated corresponding to typical workload variations in regular flight. In the medium-workload condition participants had to maintain a straight and level flight trajectory and maintain navigational parameters such as speed. In the high-workload condition participants had to simultaneously navigate around curves, adjust altitude and perform a communications task which required the memorization of call signs embedded in radio messages.

Pilot mental workload has been successfully manipulated in-flight and in simulated environments using maneuver complexity manipulations (Kiroi et al., 2016; Nittalla et al., 2018) and secondary working memory tasks (Song et al., 2011). Following a MRT framework, it is hypothesized that the increased navigational difficulty and the introduction of the call sign memorization task will place additional load on visual- and auditory-perceptual processing resources, and the working memory requirements will place additional load on verbal and spatial cognitive resources.

The predicted resources structure for responding to the two workload manipulations are illustrated in Figure 3, where the increased load on a resource is denoted by red for the navigation task, and yellow by the call sign task. Attentional resources are required to manage

visual and spatial information that is integral to a navigational task. The difficulty of the curve segments and the corresponding constant change in orientation induces an increase of psychomotor effort and visual-spatial processing. The participants must allocate attentional resources to the perception of visual-spatial information, as well as the storage of navigational parameters in working memory to enable evaluation of navigational performance. Memorizing the radio message first requires that attention be paid to the presentation of the message which involves allocating attentional resources to the perception of verbal-auditory information. To maintain the memory of the message throughout the duration of the circuit the message must be rehearsed, which involves allocating attentional resources to verbal information in working memory. Goal-oriented behavior is manipulated by two interventions: the introduction of the call sign task, and the increased navigational difficulty of the curve segment. Both manipulations will require more demand on verbal cognition for goal-maintenance (Oberauer, 2019), and therefore the verbal cognition layer of the resource structure was indicated to be relevant to both tasks. The two responses to the workload manipulation are to make navigational adjustments and to recall the radio message. Therefore, the response layer of the resource structure is indicated as manual

for the navigational manipulation, and verbal for the call sign task.

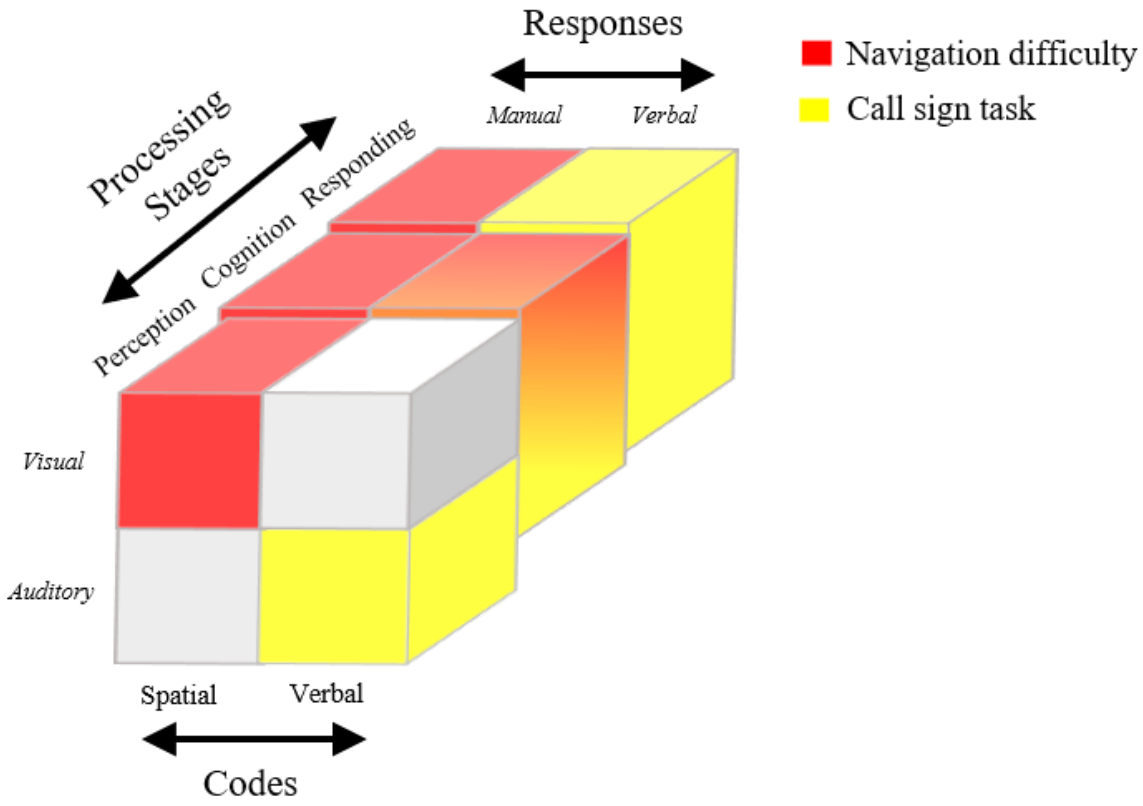

Figure 3. Wickens' (2002) Multiple Resources Theory of workload and the relation of resources and task load of the current study. The figure displays the expected structure of mental resources for the current experimental task demands. Note that the cognition level of the processing stage dimension is dichotomized only by spatial and verbal codes. The visual and auditory modalities are only distinguished in the perception level of processing, and manual and verbal responses are only relevant to the responding level.

Since the present study's workload manipulations are suspected to load visual and verbal/auditory demands, electrodes can be specified for BCI inclusion that correspond to loading visual and verbal information processing systems. Figure 4 shows electrode sites that correspond to indices that are reported to track elevated load of visual information processing, verbal information processing, and additional pertinent workload regions (e.g., central executive function at prefrontal cortex). The figure indicates which of the EMOTIV EPOC+ electrode sites are expected to be most informative to workload, and which electrodes can be removed.

Reducing the electrodes can increase the simplicity and usability of the BCI system, and also benefits classification by reducing the number of predictor variables. The figure reports findings from EEG research as opposed to other neuroimaging techniques such as fMRI for the benefit of this research project; fMRI indices often cannot translate to EEG due to volume conduction of electrical recordings. Comparing verbal forms of a Sternberg working memory task to visual and spatial versions, greater Beta power is observed at parietal and frontal-midline electrodes during the retention/rehearsal stage of the task (Hwang et al., 2005). Additionally, a working memory buffer has been suggested to be related to Beta oscillations at parietal sites (Gelastopoulos, 2019). A working memory buffer integrates perceptual information with executive demands. Integration of executive demands and perceptual information would be relevant to the current task as perceptual features of the environment must be processed to assess how one's navigational goals are being achieved. Visual information processing can most strongly be localized to parietal-occipital electrodes, which are used as an index for visual attention (Liu et al., 2016) and visual working memory load (Günseli et al., 2019). Out of the remaining electrodes the frontal-central (AF3, AF4) are most strongly implicated in mental workload studies (Herff et al., 2014; Zarjam et al., 2011) as this region of the frontal cortex is thought to be related to executive functioning (D'Esposito et al., 1995).

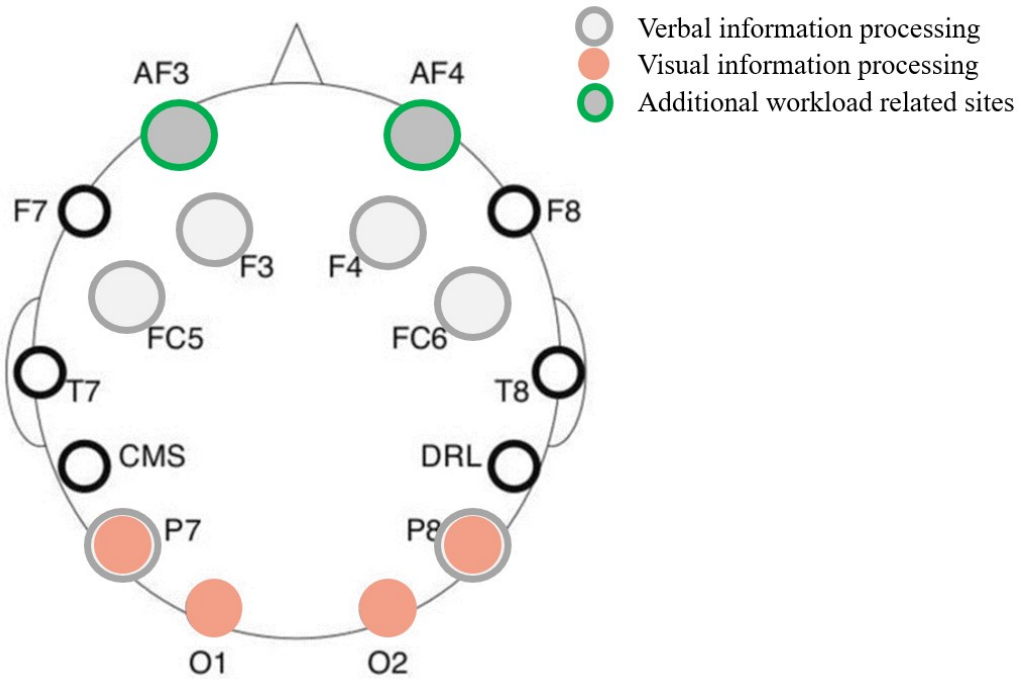

Figure 4. The expected relationship between the EMOTIV EPOC+ electrode sites and mental workload. The most reported associations between visual information processing, verbal information processing, and overall mental workload are highlighted. This figure was used to guide electrode selection for BCI classification.

The purpose of the present research was to evaluate the relationship between mental workload and neurophysiological activity during simulated flight. Intermediate changes in mental workload were selected to be representative of the variation in task difficulty during regular flight. Implementing representative changes in workload would give insight into how an EEG-BCI could perform for its desired purpose - discriminating precarious workload states from healthy workload states. Mental workload was varied by manipulating the difficulty of navigation and by having participants complete a call sign memorization task while flying. Neurophysiological activity was assessed using features of EEG that have been related to workload in other environments. EEG data was analyzed for features of the data that can discriminate between levels of mental workload and predict difficulty of flight. Two main hypotheses were examined:

- Hypothesis 1. EEG spectral features will correlate with changes in workload. EEG power densities associated with high- and medium-workload conditions will be used to establish candidate workload indices and inform spectral filtering of the BCI approach. It is expected that with increased workload, power will increase within the Beta band and decrease in Theta and Alpha bands (on average across electrodes).
- Hypothesis 2. EEG-BCI approaches will discriminate between medium and high levels of workload.

Spectral features were selected for hypothesis 1 as feature inputs to classify workload events, as they demonstrated a capacity to resist noise in similarly complex experimental conditions (e.g., Dehais et al., 2019). In comparison to other commonly used EEG features of workload, such as ERPs, features of spectral power were robust to noise sources such as movement artefacts and environmental electromagnetic interference. The frequency bands expected to be most influential are the Beta, Alpha and Theta bands. Since the engagement index appears to be quite robust, functioning in complex, multitasking environments, and with different minimal sensor arrays, features within the range of the engagement index are expected to be most informative to workload changes in the context of the present study. Therefore, when examining feature densities globally, it is hypothesized that there will be greater Beta power densities, and lower Alpha and Theta power densities with increased workload.

Although the motivation of conducting the present study was to administer comparatively mild workload changes that aligned with regular flight experience, it is expected that workload manipulations will produce detectable differences in mental state to accommodate hypothesis 2. Following an MRT rationale, it is expected that there will be overlap in the cognitive mechanisms used to manage the two workload manipulations. Using an MRT framework is

conducive to the use of intermediate manipulations that are realistic to flight, meanwhile providing sufficient contrasts in task difficulty. EEG indices of workload will be evaluated to determine which is most promising for future research in aviation safety.

2. Methods

2.1. Participants

Forty-seven participants from the university community with no flying experience were recruited for the present study. All participants were briefed on task requirements, and experiment materials before providing written consent. Ethics were approved by the Carleton University Research Ethics Board (CUREB). Participants were reimbursed for their participation with refreshments and course-credit.

As shown in Figure 5, participants were removed at various points along the experimental protocol. The final number of participants was 11. At the first stage, nine participants were removed from the dataset due to an inability to wear the EEG headset, or if the EEG malfunctioned or was incorrectly calibrated. At the second stage, nine participants were removed due to not completing the experiment. At the third stage, four participants were removed due to poor EEG quality. The automated pre-processing script (available at https://akfraser96.github.io/) removed over 80 of 240 EEG epochs for these participants. The fifth stage is detailed in the BCI Inclusion section, and the sixth stage is detailed in the BCI results section. For the final 11 participants, ages ranged from 18 to 38 (Mean = 23.6 and SD = 6.2).

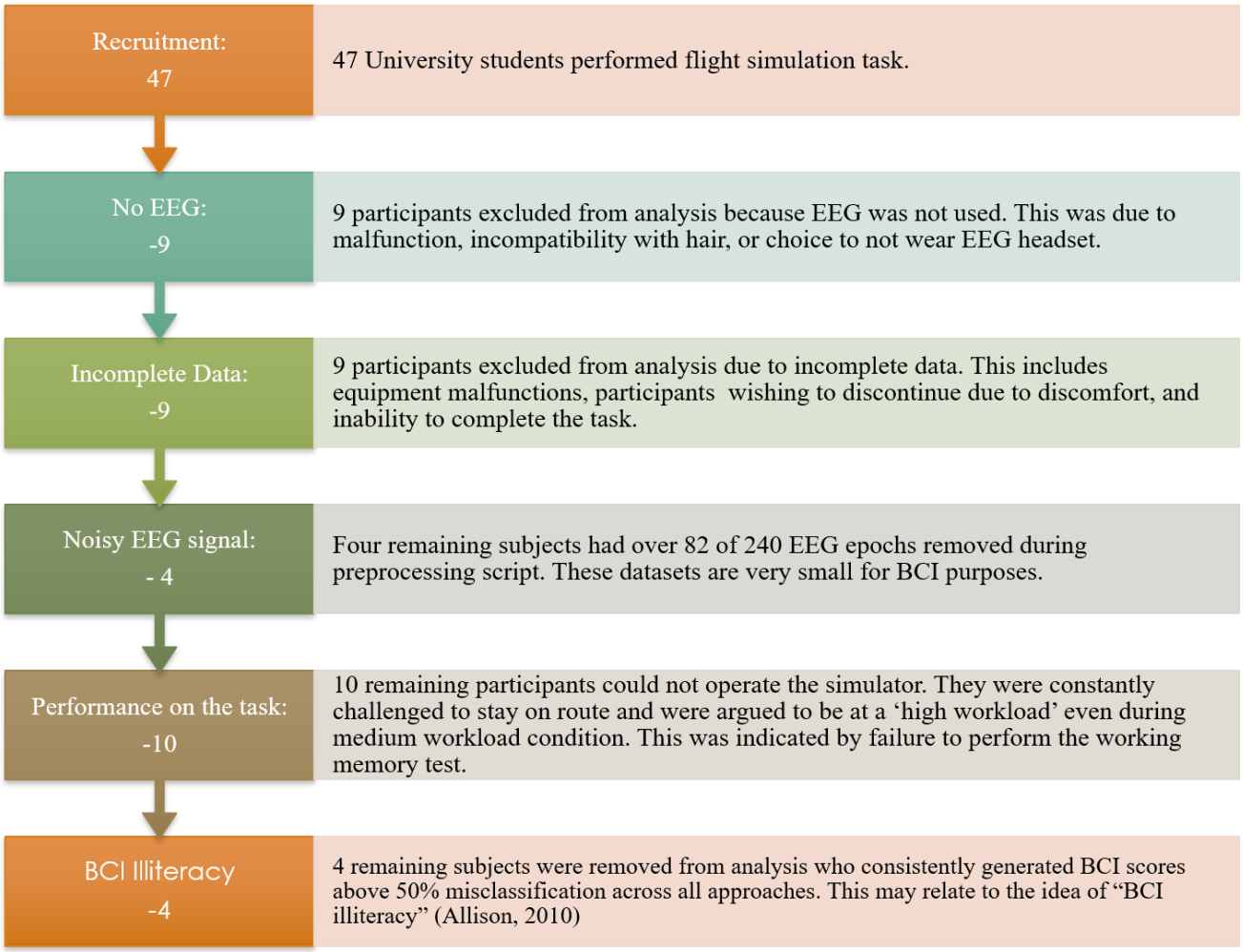

Figure 5. A chart illustrating the procedure behind participant removal. Eleven participants were deemed appropriate for BCI analysis.

2.2 Procedure

After obtaining informed consent and briefly introducing the flight simulator and the VR and EEG headset, participants were presented with a slide show that illustrated task requirements and briefed the participants on the VR flight simulator and the flight control unit. Before beginning the experimental portion of the study, the participants were provided with 20-minutes of guided practice time in the simulator, to ensure that they were able to control the aircraft and understood the tasks. Participants 'flew' simulated flight circuits at an aerodrome using VR simulation. The aerodrome was located in Hong Kong, consisting of coastal and mountainous

terrain. Participants flew three practice circuits and four test circuits. Half of the test circuits contained a radio message call sign memorization task (explained further below), which was also performed in two practice circuits for familiarity. As shown in Figure 6, for each circuit, participants were instructed to navigate through a series of large rectangular hoops which outlined the oval path of the circuit (hoop altitude ranged from 500 to 1500 feet above sea level). Circuits were initiated at altitude of the first hoop at the end of the downwind leg of the circuit (see Figure 6) and takeoffs and landings were not incorporated in the task. Participants were instructed to monitor airspeed, trying to maintain as close to 100 knots as possible on the downwind leg, as well as to manipulate the flap control (see Figure 7) at specified locations. The locations for flap adjustments corresponded to the points in the circuits where pilots would normally activate their flaps to assist with landing and then set the flaps to 0-degrees once over the runway. Each circuit took approximately six minutes to complete. After each circuit, participants were presented with questionnaires. Participants were queried about their comfort relating to the VR system and asked to recall the call signs after high-workload circuits.

2.2.1. Medium- and high-workload manipulations

High Workload. The high-workload (HWL) condition included all the flight time that occurred during the crosswind and base legs in the circuits that included the call sign tasks. During either the first and third (or second and fourth, depending on counterbalance) circuits participants were instructed to listen for and mentally rehearse the aircraft call signs mentioned in pre-recorded air-to-air communication messages (e.g., “Pendleton Traffic, this is *Delta Echo Foxtrot*, Cessna 150, Five Miles to the Northeast, Inbound for touch and gos”). To cue participants as to the start of a high-workload circuit, at circuit initiation, participants were played a message indicating that there would be radio messages played during that circuit.

Participants were instructed to remember the call signs (e.g., “Delta Echo Foxtrot”), which were consistent across three different radio messages for that circuit. After each circuit, the simulation was paused, blanked, and participants were asked questions related to comfort, engagement, and to recall call sign letters if messages were played. In addition to the call sign task, the HWL portion of the flight was limited to the crosswind and base segments of the circuit (see Figure 6). These segments of flight were expected to be challenging due to the necessary rapid adjustments in altitudes and headings.

Medium Workload. The medium-workload (MWL) condition was all flight that occurred during the runway segments in the circuits that did not contain the call sign task. This was straight flight without curves or changes in altitude, heading, or airspeed.

Confirming Workload Manipulation. The modulation of difficulty was confirmed via an analysis of participants deviation from flight path, as measured by distance from the center of target hoops. Participants’ navigational accuracy was significantly worse (16ft) around the curves than during the straight segments of flight, $t = 2.3$, $p < .02$. Additionally, 26 out of 27 total critical incidents, such as crashing, occurred during the curves, while only one occurred over the runway. Sixteen of the critical incidents occurred during trials with a call sign task. Further support for confirmation of workload manipulations were provided by performance on a peripheral detection task (PDT; see Measures section for details). Analysis of PDT shows that participants had a lower hit rate in the HWL (92%) versus the MWL (95%) condition, $t(60.9) = -1.85$, $p = .035$. There was no effect of workload on PDT response times, with both conditions resulting in an average 1.0 second button response, $t(60.5) = .86$, $p = .196$. The PDT analysis confirms that the addition of a call sign task did increase mental workload.

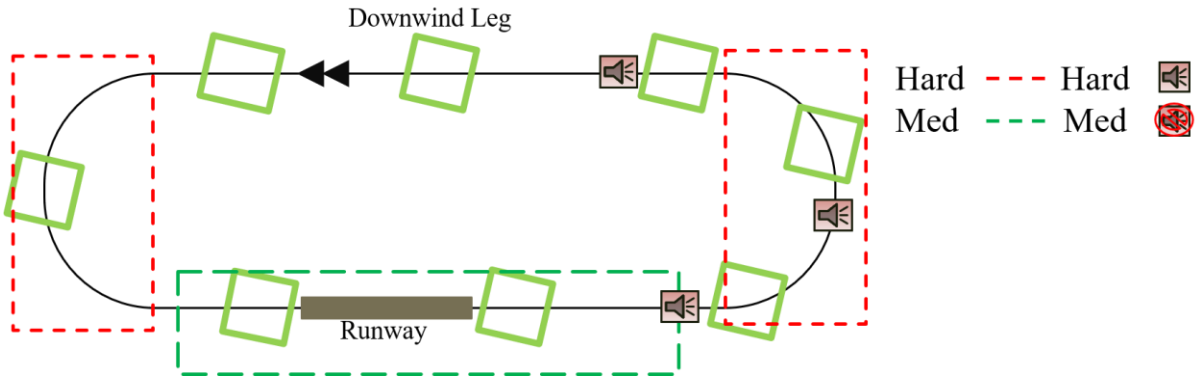

Figure 6. Illustration of flight circuit. Participants began each circuit at the location of the double arrows and at altitude of the first hoop (in green). The red dashed line outlines the navigationally challenging portions of the flight path (Base and Crosswind legs) and the green corresponds to the easiest section of flight (runway leg). Each curve took approximately 40 seconds to complete and each straight leg approximately 140 seconds. The speaker symbols represent the locations where prerecorded messages were played.

2.3. Equipment

2.3.1. Simulation environment

An HTC Vive VR headset (2016) was used to graphically display the 3D flight simulation, including the full Cessna 172 model aircraft and all exterior terrain and airspace. The flight simulation was produced by Lockheed Martin's Prepar3d software. The location was geo-specific terrain consisting of coastal and mountainous regions surrounding an aerodrome in Hong Kong. There was a 360-degree virtual environment. Flight instruments were made visible in the simulation and corresponded to the physical locations of the yoke, throttle, and flaps in the flight control unit (See Figure 7). The simulation produced aircraft realistic visuals and engine noise. Weather conditions were clear with no experience of turbulence.

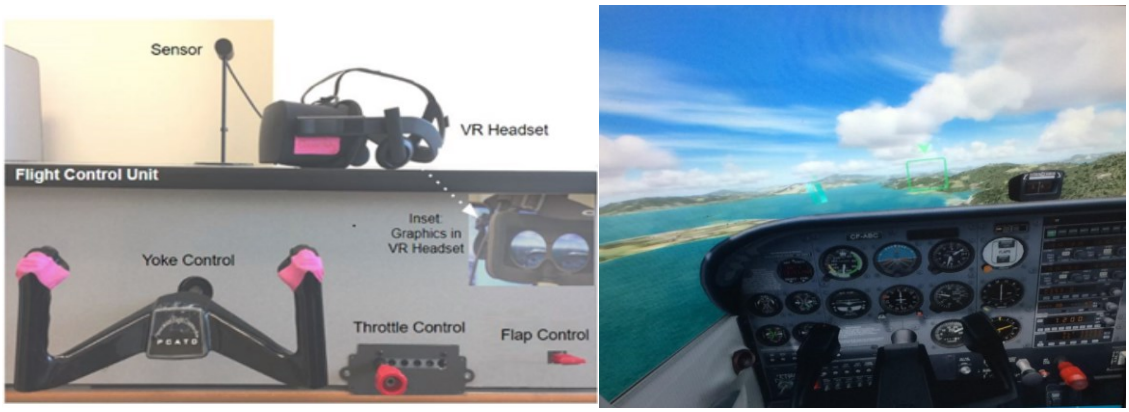

Figure 7. Display of the physical instrument layout used to perform VR flight (left) and example of participant view of the simulation (right).

2.4. Measures

Flight performance:

- **Hoop accuracy.** Hoop accuracy was measured as the absolute distance (ft) from the center of each of the eight target hoops. Absolute distance was recorded as participants intersected the horizontal position of each hoop.
- **Critical incidents.** Critical incidents were noted by the experimenters and were defined as crashes or veering too far off the flight path to recover trajectory.

Working memory load:

- **Call sign task.** After HWL circuits, participants were asked to recall the three-letter call sign that was played during the circuit. Participants received one point for each correctly identified letter of the call sign. The call sign task was performed twice, resulting in a total of six possible points. Participants who received less than three points were considered low performers.
- **Peripheral detection task.** To confirm that workload was higher in the circuits where the call sign task was present, an auditory peripheral detection task (PDT) was performed during all test circuits. A PDT is used to demonstrate workload differences when the PDT

has longer response times and/or lower hit rates during the presumed HWL conditions (Martens & Van Winsum, 1999) Tones of 1000 Hz and a duration of 100ms were played at 60 dB at pseudo-random intervals between 4-7 seconds through earbuds. Participants were instructed to press a button located by the thumb on the left-hand side of the yoke as quickly as possible upon detection of the tone. There were approximately 40 tones per circuit. The tones were played through a sound mixer to not interfere with the presentation of radio messages. During practice trials, participants confirmed that the tones did not cause interference in their perception of radio messages. Reaction times and accuracy were collected.

2.5. Electroencephalography

Electrophysiological data was collected using an EMOTIV EPOC+ 14 channel wireless EEG system with electrodes located at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. The channel placements follow the international 10-20 system and were referenced online to electrodes P3 and P4. Channels AF3 and AF4 had to be positioned underneath the top of the VR headset to accommodate the simultaneous use of the two devices (see Figure 8). The EEG recordings were collected at 2048 Hz, and then down-sampled to 256 Hz and were transmitted wirelessly via Bluetooth to an iMac desktop computer. EEG calibration involved applying saline solution to felt pads and monitoring the signal quality, assuring impedance levels remained in the 10–20 $k\,\Omega$ range. The EEG data was recorded using the EMOTIV software TestBench, applying a bandwidth of .2 to 45 Hz before further processing using the open source EEGLAB (Delorme et al., 2004) running on MATLAB v. 2019b. There was a researcher assigned to monitoring EEG activity throughout the test phase, and calibrations were performed on noisy electrodes between

circuits, taking care to keep impedance minimized and electrode placement consistent throughout the experiment.

Figure 8. The configuration of the HTC Vive VR headset and EMOTIV EPOC+ EEG headset for simultaneous use.

2.5.1. Preprocessing

During the post-collection phase, eye blink and movement artefacts were identified using Independent Component Analysis (ICA) and using an automated artefact detection algorithm (ICLabel; Pion-Tonachini et al., 2015). The automated preprocessing pipeline (see Appendix C) began with removing noisy electrode channels. Channels that had a probability distribution or kurtosis beyond 2.5 standard deviations from mean channel activity were removed. Using a blind source separation algorithm (SOBI; Belouchrani & Cichocki, 2000) periods of activity were labeled as blinks or muscle movement if they were identified at > 60% probability. Flagged

components were automatically removed, and then recordings were re-referenced, and ICA was performed a second time.

Before preprocessing each participant had 120 1-second data points for MWL and 120 1-second data points for HWL. Participants had on average 38 data points removed due to the preprocessing criteria. Samples per condition were fairly balanced after preprocessing, with an average of 2.2 fewer data points in the HWL condition (SD = 6.8). The greatest participant condition imbalance was 17 fewer samples in the HWL condition, which equates to a 9% imbalance between the two conditions sample size. Four participants had more than 80 data points removed. As per stage three in participant exclusion (Table 2), these participants were excluded from BCI analysis due to insufficient data quantity for BCI classification.

2.5.2. BCI

The goal of the BCI was to classify mental states of HWL and MWL using continuous EEG data. Frequency band features were selected, as they have been shown to be more resilient against noisy task environments such as flight tasks (Dehais et al, 2019). A possible explanation for this resilience of EEG frequency features is that their measurements are aggregated across a time interval, which attenuates the influence of an artefact, whereas temporally specific features such as an ERPs can only attenuate the artefact by averaging across measurements. A variety of BCI approaches were explored by changing combinations in feature extraction, spectral filtering, and algorithmic (machine learning) classification methods (described in following sections).

2.5.2.1. Participant inclusion

Although participants were controlled for flying experience, and had little VR experience, there was high variability in skill acquisition. Ten participants were removed because they could not sufficiently handle the simulator to perform the required tasks. A lack of

adjustment to the simulator was indicated by poor performance on the call sign task. Participants who performed worse than 50% on the call sign task were expected to be at a high level of workload during all conditions based on their difficulty with the simulator. Since the call sign task was relatively easy (remembering the letters of the call sign that was presented three times in each circuit, including within MWL segments), it was possible that failing to do so meant that mental resources were saturated by the simulator demands. BCI analysis began with 15 participants. A final analysis was performed on 11 participants, as four participants were identified as likely having incompatibilities with the BCI system.

2.5.2.2. Feature Selection

A preliminary analysis of power spectral densities (PSD) were evaluated to inform feature selection for the classification model. The difference in PSDs between HWL and MWL were calculated for each frequency band and electrode site. The frequency bands in the range of the EMOTIV EEG recording capabilities are Delta (1-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), and Beta (12-32 Hz). Figure 9 shows the first stage of examining for expected workload related effects and spatial patterns, using EEG electrode clusters. The workload effects were obtained by 1) determining each participant's median voltage level across HWL and MWL epochs. Medians were used to mitigate the influence of outliers, which likely relate to sources of noise that are ubiquitous in EEG research. 2) Each participant's median MWL voltage was subtracted from their HWL median. 3) A mean of participant workload voltage differences was calculated and presented in the figure. Red and orange cells show PSDs that increased with increased workload at a particular frequency range and electrode cluster. Blue cells show PSDs that decreased or stayed nearly the same with increased workload. Beta had a relatively greater increase in power with elevated workload in comparison to Alpha and Theta, which both had decreases in power

within some electrode clusters. The observed power density effects are consistent with the Alpha/(Beta+Theta) engagement index, which has been successful in predicting workload level against other indices (McMahan et al., 2015; Pope et al., 1995). Secondly, there is elevated Delta power across all electrode clusters, but with the largest effect at the right anterior cluster. Although there has been less research evaluating Delta frequencies, they have been reported to reflect working memory load information in frontal regions, but typically reduce in power with increased workload (Zarjam et al., 2011).

Importantly, the emergence of an engagement index pattern provides support that the conditional effect, and the EEG signal was related to workload. Given that the EEG index was evaluated in stimulating/multitasking environments such as the MAT-B and action shooter video games the engagement index is more likely to stand out in ecological settings such as simulator activity.

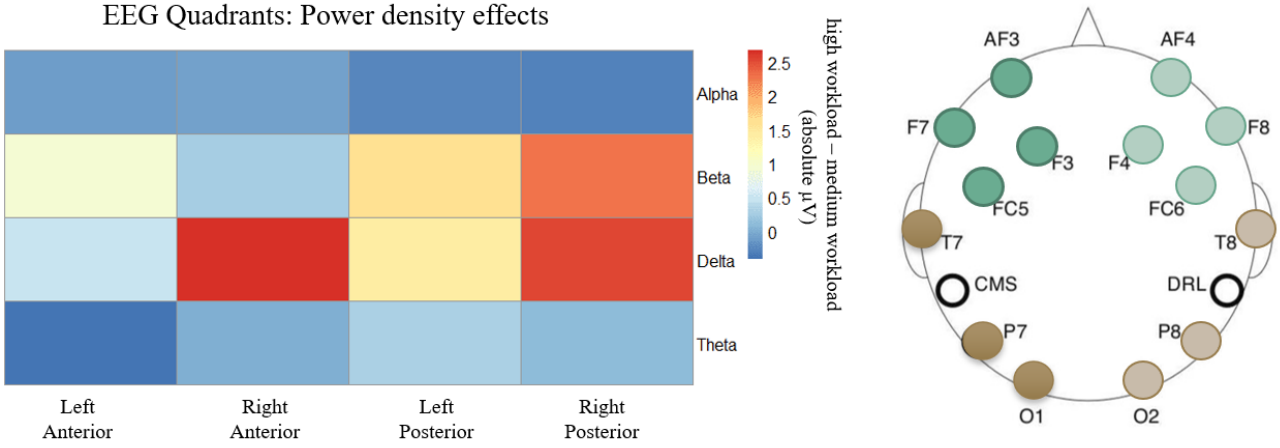

Figure 9. The results of the preliminary analysis on workload and PSDs. This figure displays the effect of workload on PSDs at each frequency band at each "EEG quadrant". Effects sizes were measured by subtracting the absolute voltage in MWL conditions from HWL. The electrodes included in the quadrants are shown on the right.

Delta activity has been associated with ocular artefacts (Tatum et al., 2011b), muscular artefacts (Stone et al., 2018) and respiratory artefacts (Yomada & Meng, 2012). Even with ideal

signal acquisition very little data would be sampled from the Delta range due to the low sampling frequency, therefore a minimal number of artefacts could cause considerable amount of noise. Therefore, the frequency range of the final model was restricted to 4-32 Hz.

2.5.2.3. Feature extraction

2.5.2.3.1. Common spatial patterns

Common spatial Patterns (CSP) is a spatial filtering technique selected for feature extraction as they optimize differentiation of band power features and are therefore recommended for oscillatory BCI uses (Lotte et al., 2015). CSP functions by maximizing the difference in variance of the measurements of different output classes. The variance of signals is contingent on the power of an oscillation, therefore separation is related to discriminating power band features. CSP has been shown to outperform other approaches such as deep neural network techniques (e.g., convolutional neural networks) when there is limited training data, which is often unavoidable in BCI experiments (Appriou et al., 2020). CSP also performs very well when there are two output classes, as in the given experiment (Grosse-Wentrup & Buss, 2008).

A spectrally-weighted variant of CSP (spec-CSP, Tomioka et al., 2006) was selected for feature extraction. Spec-CSP is reported to perform better under conditions where the parameters of an oscillatory feature are not entirely understood. Spec-CSP should be advantageous in the context of the present study, where several oscillatory patterns are expected, but there is uncertainty in how similarly those patterns present in complex flight environments in comparison to laboratory environments. Spec-CSP performs feature selection through an iterative process alternating between spatial and spectral optimization. After completion of one iteration, standardized log-variance features are obtained and then processed by the next iteration. Another feature of spec-CSP is that it enables specification of a prior. Including a prior is beneficial in

more exploratory contexts, where one can apply subjective confidence in a proposed index (e.g., elevated Beta power), but interactions with features outside the ranges of the prior are still possible. Spec-CSP has been applied successfully in determining oscillatory features of workload. Exploration of competitive techniques determined that spec-CSP performed better in the present context than regular CSP, filter bank CSP and log-bandpower estimates. Figure 10 shows the medians and distribution of classification rates per participant for each evaluated feature extraction method.

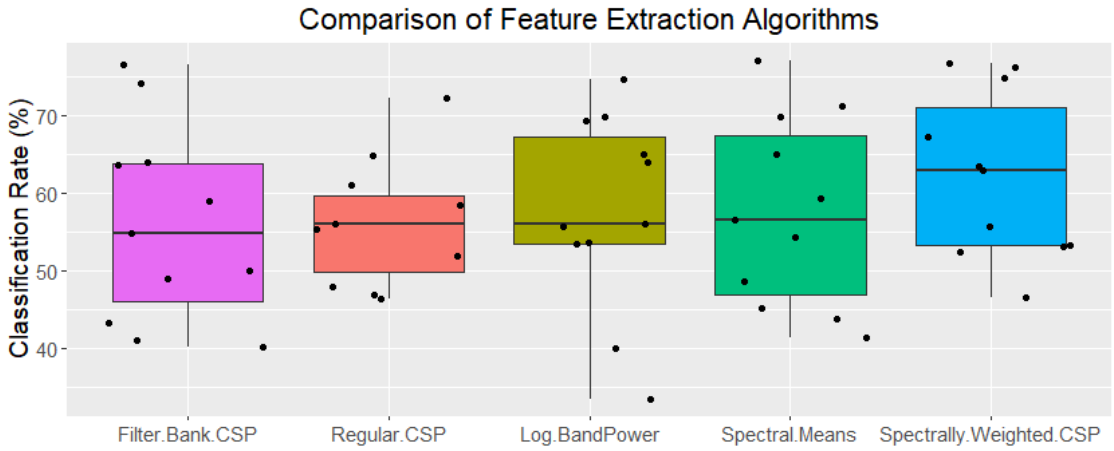

Figure 10. Distributions of evaluated feature extraction methods. All approaches used shrinkage linear discriminant analysis and a spectral filter of 4-32 Hz. Black horizontal lines indicate median values.

2.5.2.4. Classification

2.5.2.4.1. Scheme

Classification rates (CR) were used to evaluate the performance of the classification models per participant. Scores were obtained through a 'leave one out' cross-validation scheme where for every six data samples, five were used for training the model, and one was used for prediction. There were 120s of data each for the MWL and HWL conditions. High-workload samples were obtained from two 30-second windows centered on the middle of each curve for the two circuits with the call sign task. Medium-workload samples were obtained from a 60-

second window centered on the middle of the runway leg for the two circuits without the call sign task. The length of sampling windows was selected to remove the transition periods between levels of workload to avoid the potential of mislabeling events. The data were epoched into 1-second intervals to create discrete samples for the data set. One second is an effective interval for obtaining frequency band features (Nittala et al., 2018). The sampling method resulted in a balanced class 240-point data set, where 200 points were used to train the model, and 40 used for testing.

2.5.2.4.2. Shrinkage linear discriminant analysis (sLDA)

Linear discriminant analysis uses class mean and covariance matrices to estimate a decision boundary for assigning inputs to a response class. The decision boundary is a hyperplane that maximizes the difference between class means and minimizes the interclass variance. Covariance matrices can be a poor estimator when the number of training samples is small compared to the number of features (Lotte et al., 2007). When shrinkage is applied the standard empirical covariance matrix is replaced by a diagonal matrix of variance estimate, which provides more reliable estimates of decision boundary points with sparse amounts of data (Binias et al., 2018). Lotte et al.'s review of classification techniques for EEG-BCI recommended use of sLDA in part due to the fact that EEG-BCI often have practical limitations which results in non-optimal quantities of data. It is recommended to have at least five to ten times the amount of data per outcome class than the number of dimensions (Jain & Chandrasekaran, 1982; Raudys & Jain, 1991). The present studies dimensions are three frequency band ranges * 14 electrode sites = 42 and the sample size is 120 per class and does not meet aforementioned requirements. Therefore, sLDA may be effective due to eliminating the cost of certain data impracticalities.

Shrinkage LDA has been successfully implemented in EEG-BCI workload classification in-flight (Dehais et al., 2019). Dehais et al., 2019 implementation of sLDA resulted in a classification rate of 71% for high- and low-workload periods. The performance of the classifier was arguably quite successful as classification was achieved despite signal acquisition complications resulting from electromagnetic properties of an aircraft. Although spec-CSP performs well with other classifiers such as logistic regression, sLDA marginally outperformed other evaluated classifiers with use of spec-CSP. Figure 11 shows the medians and participant classification rate distributions for the explored machine learning algorithms.

2.5.2.4.3. Quadratic discriminant analysis (QDA)

Instead of a linear decision boundary between class predictions, QDA develops quadratic decision boundaries. Oscillatory patterns of neural activity are thought to arise from excitatory neuronal assemblies that are non-linear (Setareh et al., 2018). Therefore, a quadratic function may be expected to best model the EEG features of workload. Quadratic discriminant analysis may also be a promising alternative to LDA, as a main limitation of LDA is that its linearity can cause it to perform poorly when nonlinear EEG data is particularly complex (Garcia et al., 2003).

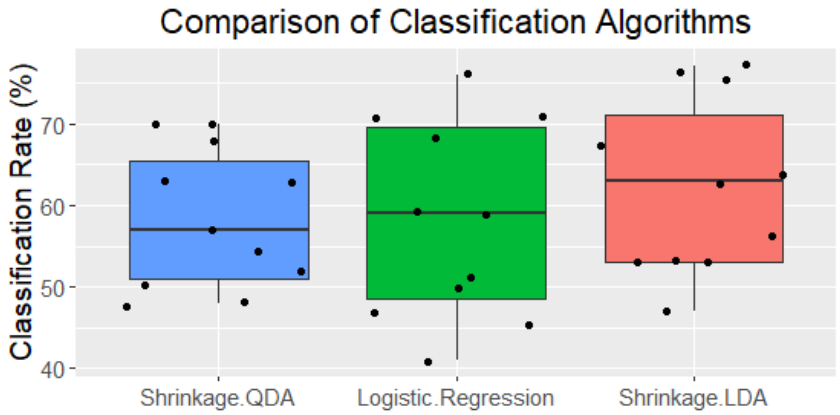

Figure 11. Comparison of classifier performance under a spec-CSP paradigm and spectral filtering of 4-32 Hz. Horizontal black lines indicate median values.

2.3.3.4.4. Temporal filtering

An additional analysis was performed to evaluate the effect of classification on a more temporally proximate dataset. Temporally 'binning' circuits into classification sets should minimize the potential impact of temporal artefacts, such as change in impedance, unintentional shifts in electrode positions, and change of cognitive strategies. Shifting between HWL and MWL conditions may illicit fatigue, and several participants reported minor increases in VR related discomfort, such as nausea. Factors of VR comfort and the demand of the simulator may have the potential to impact mental state and related EEG signals in later circuits. The analysis examined the first two circuits and final two circuits separately and obtained final classification scores by averaging the classification scores of both halves.

The analysis of BCI performance began with evaluation of spectral filtering approach to determining the optimal frequency range for the model. The performance of the model was evaluated on predictive power and minimization of features. An analysis of different data segment comparisons was performed on the selected model to address temporal dependencies. Once the optimal strategy was determined that minimized temporal non-cognitive effects, evaluation of error types was performed.

3. Results

3.1. Best Performing EEG Spectral Frequency Features

Figure 12 shows the relative classification performance of different spectral filtering approaches. All BCI approaches were performed using a spec-CSP for feature extraction and sLDA classifier as detailed in the BCI methods section. Classification rates are plotted from left to right based on best to worst scores per spectral filtering approach. The performance per frequency band from best to worst is: Beta, Alpha, Theta, Delta. Delta was the worst performing

frequency range and was removed from the final model, as it increased the error rate. The selected spectral filtering method was 4-32 Hz.

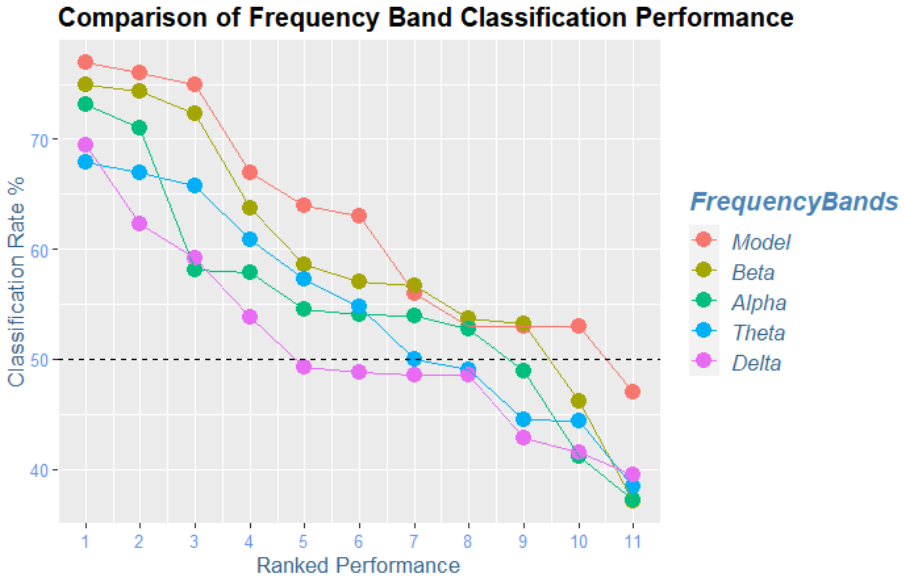

Figure 12. Comparison of classification performance of different spectral filtering approaches. The full model (red) contains oscillatory information between 4 and 32 Hz. Scores from left to right are ordered from better to worse performance separately for each approach (i.e., participant order is varied for each line graph, and y columns do not necessarily correspond to the same participant).

3.2. Classification

3.2.1 Classification approaches

Classification approaches were created selecting different electrode combinations as guided by electrode indices of workload (displayed in Figure 13). Figure 13 shows the classification distributions across participants for different approaches. First all electrodes were included which resulted in a mean classification rate of 56.5% (SD = 13.5%). Classification was improved to a mean of 61.4% (SD = 11.5) with electrode reduction to the identified index electrodes (AF3, AF4, F3, F4, FC5, FC6, P7, P8, O1, & O2). The next reduction involved identifying index sites that may be corrupted by noise; particularly those that are sensitive to blinks or dynamic visual input. Removing occipital electrodes (O1 & O2) increased the mean

classification rate to 63% (SD = 11.9) which was better than removing both occipital and frontal (AF3, AF4) electrodes (mean = 58.9%).

After evaluating several approaches, modulating the feature extraction process and spectral filters, there were four participants who performed consistently worse than chance. The poor BCI performance for select participants could be related to the phenomena of "BCI illiteracy" which refers to individuals who cannot successfully control and integrate into a BCI system. BCI illiteracy is suspected to occur in about 20% of participants and is suggested to reflect measuring limitations from individual neuroanatomical differences (Allison & Neuper, 2010), which is problematic for both active and passive BCI. Regardless of performance and adherence to the task, some individuals' neuroanatomies are not conducive to detecting the relevant signals. In the present context, a BCI measuring limitation could likely be explained by implementing an EEG headset that could not be customized to account for individual differences in head size and shape. A Welch two-sample t-test confirmed the difference between good and bad performers were significant, $t(12.55) = 6.46$, $p < 0.001$. The four possible BCI illiterate participants were removed from BCI analysis and are highlighted in Table 2. The final reduction was in removing the four identified BCI illiterate subjects from classification using only the index electrodes minus occipital sites. Removing the four BCI illiterate subjects increased classification rate to a mean of 68.3% (SD = 7.9%).

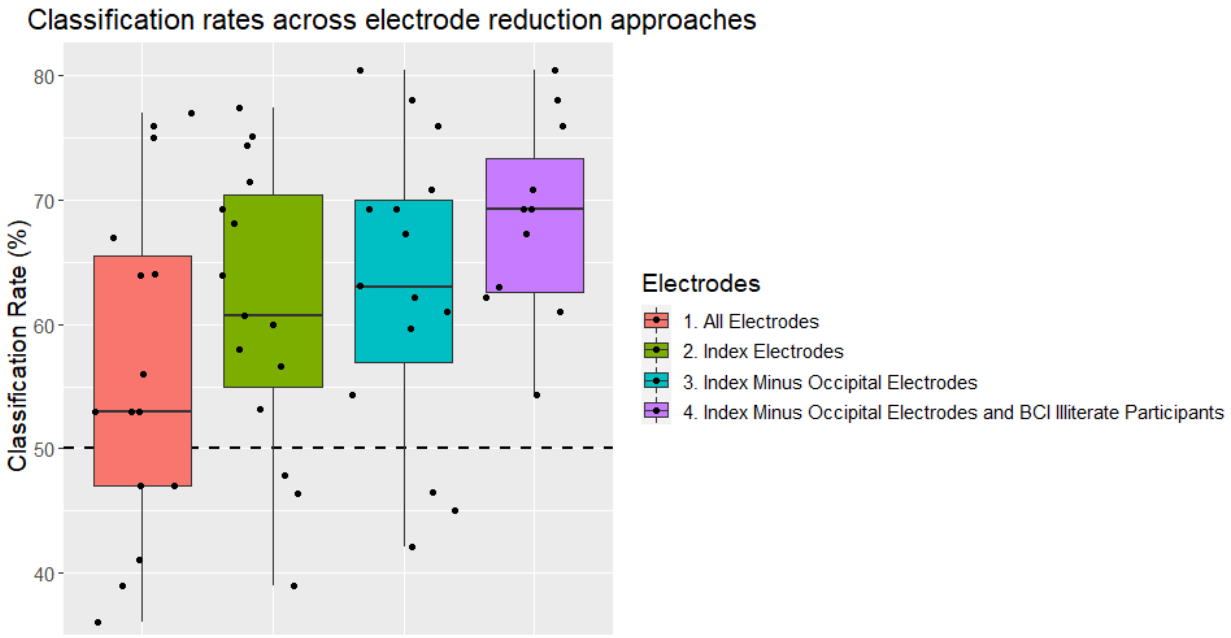

Figure 13. Distributions of participant classification rates as electrode selection was refined. The 'index electrodes' are described in Figure 4. The occipital electrodes were identified as potential sources of noise as they are sensitive to blinks and modulating visual inputs. BCI 'illiteracy' occurs in about 20% of subjects where classification cannot occur, likely due to neuroanatomical properties (Allison & Neuper, 2010).

Table 2

Participant Performance Across a Variety of Approaches

Participant	BP.Wide.QDA	CSP.Wide.QDA	BP.Wide.LogReg	CSP.Wide.LogReg	BP.Wide.COV	CSP.Wide.GLM	sCSP.BETA.QDA	sCSP.WIDE.LDA	sCSP.WIDE.LogReg	sCSP.WIDE.QDA	sCSP.LDA.theta	sCSP.LDA.beta	sCSP.LDA.alpha
2	57	42	32	31	53	30	38	64	59	57	68	46	49
15	73	71	70	73	72	71	69	77	76	70	67	72	71
17	34	39	28	31	48	31	48	56	50	54	55	54	41
18	51	37	32	31	55	30	46	41	32	33	41	48	48
19	58	63	55	46	53	45	65	67	68	63	57	74	58
20	29	24	33	26	30	27	25	36	36	35	43	33	41
21	71	59	60	66	72	64	72	75	71	70	44	75	37
22	43	52	54	47	82	47	60	63	59	63	44	57	53
24	63	45	51	47	65	47	51	53	45	50	38	57	58
32	62	42	43	48	66	46	49	53	47	52	61	53	54
33	45	33	31	31	43	29	38	47	41	48	50	37	54
34	64	52	51	60	61	59	60	76	71	68	66	64	73
37	37	22	24	26	41	24	37	39	33	44	50	35	51
39	36	29	33	32	39	30	49	47	41	25	41	56	29
42	60	38	36	48	55	47	54	53	51	48	49	59	55
Approach mean	**52**	**43**	**43**	**44**	**56**	**43**	**52**	**56**	**51**	**52**	**51**	**55**	**52**
Approach median	**55**	**40**	**40**	**46**	**55**	**45**	**50**	**53**	**49**	**51**	**50**	**56**	**53**
Mean, Exluded PPTs	**57**	**49**	**46**	**48**	**61**	**47**	**55**	**62**	**58**	**59**	**55**	**59**	**55**
Median, Exluded PPTs	**60**	**45**	**51**	**47**	**61**	**47**	**54**	**63**	**59**	**57**	**55**	**57**	**54**

Note. Approaches were altered through different combinations of feature extraction, classifier, and spectral filtering methods. The four highlighted participants exhibited consistent poor performance (< 50% classification rates) and are implicated as potential 'BCI illiterate' candidates. BP = logarithmic band-pass, CSP = common spatial patterns, sCSP = spectrally-weighted common spatial patterns, Wide = 4-32 Hz spectral filtering.

3.2.1.1. Temporal analysis.

Figure 14 shows the classification distributions under different temporal schemes. The first distribution shows the classifications using oscillatory features as predictors and the first and last circuit as the output classes. There is a large effect using time as the output class, indicating time in the experiment could have influenced classification. A one-sample t-test indicates that the first and final circuits can be classified significantly more accurate than chance, $t(10) = 6.83$, $p < 0.001$, without explicit reference to cognitive manipulations. Additionally, four out of 11 participants' classification rates were above 96%, meaning temporal effects, or sub-optimal calibration properties (e.g., impedance, electrode displacement) were highly influential. The temporal effects observed in the present study could be related to adoption of cognitive strategies, fatigue, or alterations in properties of signal quality, such as impedance. When evaluating the effect of workload manipulation in the first two circuits (proximate to calibration) classification rates were higher than original cross-circuit classification, $t(18.5) = 3.25$, $p = 0.004$ (Figure 14, second boxplot) and classification of final 2 circuits, $t(19.7) = 2.23$, $p = 0.04$. A decrease in classification accuracy at later circuits is in accordance with a short calibration fidelity where signal acquisition degrades over time from EEG set-up.

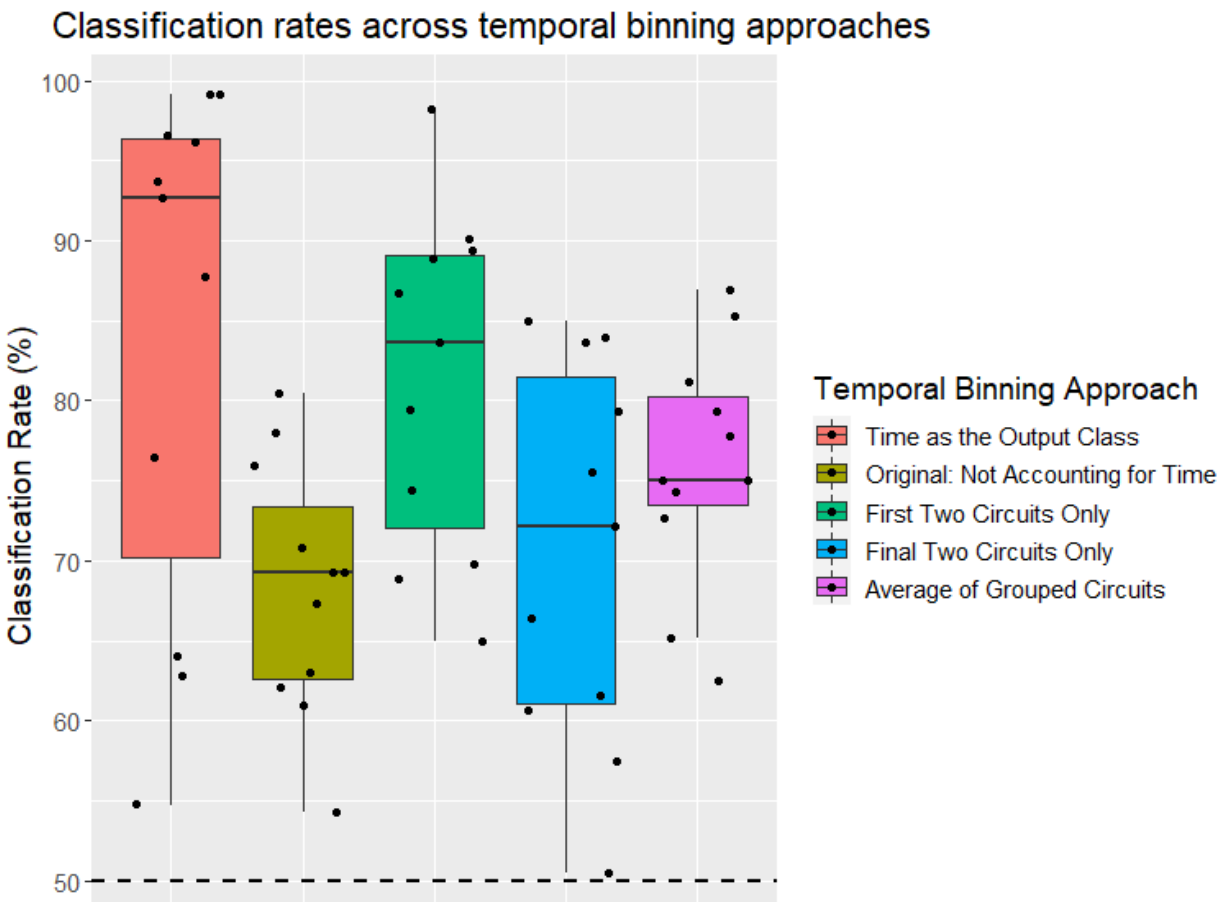

Figure 14. Classification rates for several different data segments. 'Time as Output - Class' consists of data from only the first and last flight circuit. Horizontal black lines indicate median values.

Classification rates for the first two circuits and last two circuits were averaged for each participant to reduce temporal artefacts. The mean classification rate from averaging across consecutive circuits was 75.9% (SD = 7.5%) which is 8% better than classifying across all circuits without temporal binning, $t(19.9) = 2.32$, $p = 0.03$.

3.2.2. False Negatives

False negatives are another important metric for classifier performance, as falsely warning of periods of mental overload presumably carry little risk, in comparison to failing to detect them. Failing to detect instances of HWL could either enable existing concerns of pilots

proceeding without awareness of mental overload, or the negative readings of an instrument for HWL detection would provide a 'false security' and bias introspective assessment. Therefore, it is critical that false negative rates are low when evaluating BCI performance.

Figure 15 shows the distribution of participant error rates. The reported error rates were achieved through averaging each participants' rates from the first two circuits and last two circuits. The mean false negative rate was 22.8%, which was less than the false positive rate (26.3%), but not to a statistically significant extent, $t(16.4) = 0.86$, $p = 0.4$. The measures of classifier performance are shown in Figure 16.

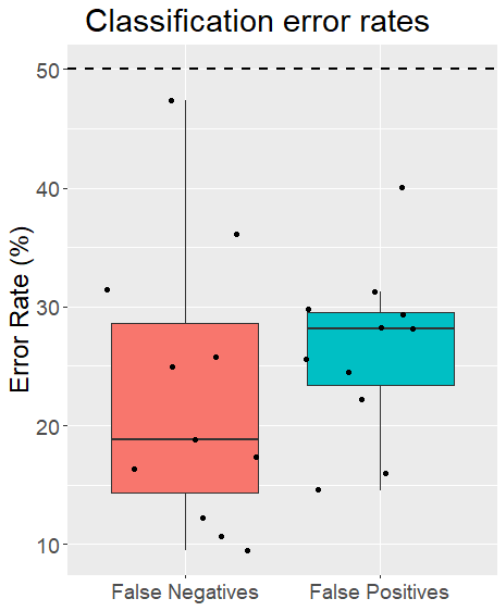

Figure 15. Distributions of participant error rates. Horizontal black lines indicate median values.

		Actual Values
Predicted Values	**High Workload**	**Medium Workload**
High Workload	Correct Identification of HWL: 64.3%	False Positive rate: 26.3%
Medium Workload	**False Negative rate: 22.8%**	Correct Identification of MWL: 73.7%

Figure 16. Confusion matrix displaying measures of classifier performance.

4. Discussion

The present research investigated an EEG-BCI for monitoring mental workload during VR flight simulation. Participants performed simulated flight operations while workload was manipulated by varying navigational difficulty and performing communication tasks. The workload manipulations were selected to increase ecological validity by corresponding with workload variations experienced in regular flight. EEG data was collected and used to classify periods of flight as high or medium workload. A classification rate of 75.9% was obtained which provides promise for future use of EEG-BCI in aviation practice.

4.1. EEG spectral features and workload

Candidate neural indices of workload were established by obtaining PSD contrasts of high and medium workload. Four frequency bands were evaluated. Of these, the Beta band features most strongly correlated to workload, and greater Beta power densities were observed with increased workload in the right and left anterior and right and left posterior EEG electrode clusters (see Figure 9). Alpha and Theta frequency bands were also informative, and displayed decreased power densities with increased workload, but only at specific electrode channels. Workload power density contrasts were also observed in Delta frequencies, particularly over the right hemisphere. However, Delta was not included in the model due to identified noise, and negative impact on classifier performance. The patterns are in accordance with the engagement index (Pope et al., 1995) and the current studies hypotheses relating to the Alpha, Beta and Theta frequency bands. The presence of the engagement index patterns of activity in the current study

suggests that signals related to increased cognitive load are robust enough to be observed in dynamic events with moderate changes in mental workload.

Following an MRT framework, it was predicted that the workload manipulations would load visual and auditory perceptual resources and verbal and spatial cognitive resources. To this end, the largest effects of workload on Alpha and Beta frequencies occurred at electrode locations that have been related to increased spatial and verbal cognitive load. The largest effect of workload on Alpha occurred as decreased power at parietal sites. The observed decrease in parietal Alpha power is in accordance with Radüntz's (2017) Dual Frequency Head Maps (DFHM) model for EEG workload indexing, which postulates a specific role of parietal Alpha. In a follow-on study, Radüntz (2020) observed a particular sensitivity of parietal Alpha to planning during increasingly complex puzzle solving. The demand on strategizing would be especially true given the structure of the current study's high-workload condition. Non-pilot participants were tasked with navigating to a distant target around a curve. Flight path deviations determined that navigating curves was a challenging task, therefore strategizing would be critical for projecting a flight trajectory, and to update strategies for future attempts. Since strategizing is a verbal process, it is possible that parietal Alpha was an index for verbal cognitive load in the present study. The largest effects of workload on Beta in the present study were increases in power at parietal and temporal electrodes.

As shown in Figure 3, the HWL condition taxed multiple mental resources associated with detecting and processing verbal information associated with the radio calls. Maintaining this information for later recall is generally referred to as working memory. New computational models are hypothesizing a role of parietal Beta activity in an episodic working memory buffer (Gelastopoulos et al., 2019). An episodic working memory buffer integrates multi-modal inputs

and relates them to long-term memory, such as procedural knowledge and prospective memory. Therefore, engagement of an episodic buffer would be critical to performance in the present study, as perceptual load was systematically varied and was pertinent to participant goals. It is possible then that parietal Beta relates to increased perceptual load and working memory (cognitive) resources.

Although Beta and Alpha power were indicative of changes in workload, it is not clear if the present study determined all the key EEG features of workload. It is difficult to assess how much relative relevant information each band contains since their fundamental properties dictate that they are sampled unevenly. For example, Beta may be most predictive to workload solely based on the fact that Beta waves were sampled more frequently. Additionally, it cannot be confirmed that Delta waves do not reflect workload information as suggested by Zarjam et al. (2011), as the current study design was not conducive to examining Delta activity. The current BCI model was trained on 1-second data points, which would contain less than four Delta signals. A sample of < four may be too small to detect discriminatory features. Additionally, Delta activity is often mistaken for artefacts originating from the jaw and neck (Nicolas-Alonso, & Gomez-Gil, 2012), which may be of particular concern in the present study as participants were not discouraged from performing regular head movements. Movement artefacts may also be related to why the classification error decreased when Delta frequencies were excluded from the model. It may also be of note that conducting studies that place more emphasis on high frequency signals may be of benefit. If determining that low frequency signals do contain workload information, it may be hard to utilize them in urgent contexts such as mental overload detection.

Less distributed workload indices such as frontal Theta were not observed in the present study (see Appendix B, Figure 18). Differences in power densities between HWL and MWL were insignificant for all frontal electrodes. One possible explanation is that frontal Theta modulations cannot withstand the additional noise from operational contexts (e.g., blinks, movement, intermediate workload manipulations), which would support arguments against using reductionist strategies to assemble interventions for 'real-world' contexts (Hasson et al., 2020). An alternative explanation is that measurements at frontal electrodes were hindered by the placement of the VR headset, which may have caused sensor displacement or impaired signal quality. The rejected EEG channels during preprocessing were predominantly frontal electrodes, but the source of noise was not evaluated.

4.2. EEG-BCI approaches and workload

The present study achieved a classification rate of 75.9%. Classification results were achieved through reducing the electrode set, reducing temporal noise, and removing participants who were poor BCI performers.

In the preliminary stages of analysis, Alpha, Beta and Theta frequencies were identified as the spectral ranges of interest, and shrinkage LDA was identified as the optimal classification algorithm for the current BCI paradigm. In the next stage of analysis, the electrodes were reduced to reflect regions of interest as reflected in the EEG workload literature. The selected electrodes included frontal, parietal and occipital regions. Classification was further enhanced by removal of occipital electrodes due to concern of signal interference. An electrode selection of AF3, AF4, F3, F4, FC5, FC6, P7 and P8 was the optimal electrode scheme and improved classification rates by 6.5%. The four worst performing subjects were removed due to their identification and potential BCI 'illiteracy', which improved classification accuracy by another

5.3%. Lastly, to reduce the large effect observed by time within the study, a temporal binning strategy was implemented which enhanced classification by another 7.6% to result in 75.9% accuracy.

Although the classification rate of 75.9% is already promising for application, there is reason to believe that the potential for EEG-BCI workload detection is greater. If the following limitations are not overestimated, then the obtained classifications should be a conservative low-end approximation of how much neurophysiological signal can be detected with EEG and correlated to mental workload. Given the ecological complexity and areas for improvement, the classification results can be considered meaningful in terms of providing promise for future use in training, and reliable monitoring during flight.

The primary consideration is that signal acquisition properties can be enhanced. The results of the current study show that classification was significantly better in the first two circuits than the final two circuits. A plausible explanation of the temporal effect on classification is that the EEG signal quality worsens from the time of calibration. For example, the sensors may get displaced or the conductive gel may dry up. BCI customization would also be an effective approach to achieving greater signal acquisition. For example, anatomical features vary from person to person and have an impact on how well they can interact with a BCI (Allison & Neuper, 2010). Customizing sensor locations would therefore be one approach to improving BCI relative to the present study.

Additionally, when evaluating classification performance through temporal binning of consecutive circuits, classification rates were significantly higher in the first two circuits than the final two circuits. The presented classification result was calculated by averaging the first two circuits and final two circuits but was arguably a conservative approach. Three plausible

explanations exist for this temporal effect. 1) Cognitive strategies changed over the duration of the study, 2) participants became disengaged through boredom or fatigue near the end of the study, and 3) EEG signal acquisition degraded over the course of the experiment. Although 3 is most plausible, I could not discount 1 and 2 with confidence. However, there was not an improvement in flight performance between the first two circuits and final two circuits, $t(10)=0.26$, $p = 0.8$, which would suggest if cognitive strategies changed (*1*), they did not become more efficient. Without increasing efficiency, different cognitive strategies should rely on similar amounts of mental resources and be just as detectable with BCI. Additionally, subjective reports indicate that participants rated the experience as immersive and engaging, and therefore it was unlikely that the task caused boredom as per explanation 2. It is also unlikely that participants became fatigued. After each ~six-minute circuit participants had roughly five minutes of rest and were encouraged to relax and enjoy some water. However, there is good evidence for explanation 3. EEG signal acquisition is sensitive to time. For example, the conductive gel that enhances signal quality dries up, the sensors get displaced with head movements, and individuals' patterns of neural activity vary over time (Lotte et al., 2007). Therefore, if strategies are executed to reduce the issue of signal acquisition, for example to recalibrate after each circuit, a classification rate closer to what was observed in the first two circuits (81.3%) might be achieved.

A recommendation for future studies evaluating physiological response to intermediate workload changes would be to control for skill. Although participants were recruited who specifically had no flying experience, large performance disparities occurred due to learning propensities. Matching participant skill level for small performance disparities would solve a large part of the present study's participant inclusion problem. Closely matched participants

would also be beneficial to evaluating small effects related to intermediate manipulations, as specific workload manipulations are strongly related to skill level. For example, the present study observed that lower-skilled participants did not have the spare mental capacity to engage with the radio message working memory task. Performance disparities could be mitigated by recruiting pilots, who would already have stable navigational control. Having pilot participants would also minimize time needed for practice, which would allow for more time to be allocated to the test phase, recalibrations, and data collection, which would address another common limitation of small samples in BCI research.

4.3. Theoretical limits

Although BCI approaches focus on predictive rather than explanatory power, theoretical work may be a useful aid to understanding the potential applications of BCI. EEG oscillations have been used to predict mental commands in BCI, however little is known about their relationship to the cognitive phenomena to which they are associated (Cohen, 2017). Buzsaki & Draguhn (2004), suggests that oscillations are an integral function of efficient activity of neuronal assemblies, as oscillations enable phenomena such as entrainment of a large network of neurons, and synchrony of converging inputs. If EEG oscillatory features are more directly related to the phenomena they are used to investigate, rather than epiphenomenal, then one should expect greater statistical means for BCI research.

Furthermore, it is not understood how perfectly mapped a particular oscillation is to a specific computation, function, or neuronal assembly (Cohen, 2017). That is, elevated Beta power could result from increased workload in the parietal region but reflect another function all together in another brain region. The inverse suggests that ratio indices such as the engagement index may not be direct measures of workload, but rather separate measures of relevant

components of task demand, and would thus be a highly operationalized use of workload to specifics of the task. The presence of an actual unitary workload mechanism may be doubtful as classifying workload level across different tasks is often unsuccessful (e.g., Baldwin & Penaranda, 2012; Walter et al., 2013). Therefore, emphasis should be given to establishing the components of workload that are pertinent to the operational context, and the resource structures required for that task.

4.4. Future Work

Future work should be aimed at dissociating the workload processing resources across particular flight procedures. Understanding which combinations of concurrent tasks cause the most interference and strain on resources is a promising way to improve flight safety. Determining load inducing task combinations could be accomplished through an array of task pairings and evaluation of performance decrements and physiological monitoring, and would enable better prediction of high-risk flight conditions. Using physiological monitoring approaches such as EEG may be advantageous in determining the relationships between particular flight procedures and separate mental resources, as these relationships may be difficult to probe through retrospective subjective reports. Determining how well EEG can detect the individual components of the resource structure and differentiate between dual task and single task dependence on a particular component could be an asset in evaluating workload in real practice.

Instead of seeking a unified model of workload, following Multiple Resource Theory, determining workload indices relating to separate processing structures may be more promising for performance monitoring given the multitasking environment of flight. Additionally, minimizing the feature space to individual components, such as motor control signals during

motion control periods of flight, may disentangle the signal from some of the other competing cognitive functions enough to make inferences about the content of the observed signals. In the current environment increased Beta amplitude could be associated with multiple cognitive tasks limiting the inferences that can be made between region and function. If an experiment was structured that one task was reliably prioritized a stronger connection can be made between function and EEG feature.

Understanding the relationship between workload and the features of the task would enable task specific interventions and pilot training. Ambitions should be toward determining specific measures pointing at where an operator is struggling (e.g., decision making, attentional engagement), and interventions can therefore also be specific. Being able to indicate the specific task challenges may also be beneficial for training. Future studies can systematically vary the demand of subtasks while maintaining a multitasking environment relevant to realistic flight.

4.5. Conclusion

In conclusion, the present study determined that there is potential for using EEG oscillatory signals for monitoring pilot mental workload in realistic flight environments. Since promising EEG-BCI results were achieved in an environment as complex as maneuvering a simulated aircraft, it is likely that EEG-BCI has potential in other dynamic operator settings, such as air traffic control. The current study also determined key brain regions and EEG oscillatory features that may be critical to identifying HWL during flight.

5. References

Abreu, R., Leal, A., & Figueiredo, P. (2018). EEG-Informed fMRI: A Review of Data Analysis Methods. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00029

Allison, B., & Neuper, C. (2010). Could anyone use a BCI? In *Brain-Computer Interfaces: Human-Computer Interaction Series* (pp. 35–54). https://doi.org/10.1007/978-1-84996-272-8_3

Appriou, A., Cichocki, A., & Lotte, F. (2020). Modern Machine-Learning Algorithms: For Classifying Cognitive and Affective States From Electroencephalography Signals. *IEEE Systems, Man, and Cybernetics Magazine*, *6*(3), 29-38.

Baddeley, A.D. (1986). Working memory. Oxford, UK: Clarendon.

Baldwin, C. L., & Penaranda, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within-and cross-task workload classification. *NeuroImage*, *59*(1), 48-56.

Belouchrani, A., & Cichocki, A. (2000). Robust whitening procedure in blind source separation context. *Electronics letters*, *36*(24), 2050-2051.

Binias, B., Myszor, D., & Cyran, K. A. (2018, April 10). *A Machine Learning Approach to the Detection of Pilot's Reaction to Unexpected Events Based on EEG Signals* [Research Article]. Computational Intelligence and Neuroscience; Hindawi. https://doi.org/10.1155/2018/2703513

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003

Bruckmaier, M., Tachtsidis, I., Phan, P., & Lavie, N. (2020). Attention and capacity limits in perception: A cellular metabolism account. *Journal of Neuroscience,* 40(35), 6801–6811.

Chen, Y., & Huang, X. (2016). Modulation of alpha and beta oscillations during an n-back task with varying temporal memory load. *Frontiers in psychology*, *6*, 2031.

Clarke, DD., & Sokoloff, L. (1999). Regulation of cerebral metabolic rate. In: Siegel GJ, Agranoff BW, Albers RW (eds) Basic neurochemistry: molecular, cellular and medical aspects, 6th edn. Lippincott-Raven, Philadelphia

Cohen, M. X. (2017). Where does EEG come from and what does it mean?. *Trends in neurosciences*, *40*(4), 208-218.

Comstock, JR., & Arnegard, RJ. (1992). The Multi-Attribute Task Battery for human operator workload and strategic behavior research (NASA TM-104174)

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Dasari, D., Shou, G., & Ding, L. (2017). ICA-Derived EEG Correlates to Mental Fatigue, Effort, and Workload in a Realistically Simulated Air Traffic Control Task. *Frontiers in neuroscience*, *11*, 297. https://doi.org/10.3389/fnins.2017.00297

Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., & Lotte, F. (2019). Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions. *Sensors (Basel, Switzerland)*, *19*(6). https://doi.org/10.3390/s19061324

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature*, *378*(6554), 279–281. https://doi.org/10.1038/378279a0

Federal Aviation Administration. (n.d.). *Fact Sheet – General Aviation Safety* [Template]. Retrieved November 17, 2020, from https://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=21274

Garcia, G. N., Ebrahimi, T., & Vesin, J.-M. (2003). Support vector EEG classification in the Fourier and time-frequency correlation domains. *First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings.* https://doi.org/10.1109/CNE.2003.1196897

Gelastopoulos, A., Whittington, M. A., & Kopell, N. J. (2019). Parietal low beta rhythm provides a dynamical substrate for a working memory buffer. *Proceedings of the National Academy of Sciences*, *116*(33), 16613–16620. https://doi.org/10.1073/pnas.1902305116

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors*, *40*(1), 79–91. https://doi.org/10.1518/001872098779480578

Grosse-Wentrup, M., & Buss, M. (2008). Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction. *IEEE Transactions on Biomedical Engineering*, *55*(8), 1991–2000. https://doi.org/10.1109/TBME.2008.921154

Günseli, E., Fahrenfort, J. J., van Moorselaar, D., Daoultzis, K. C., Meeter, M., & Olivers, C. N. L. (2019). EEG dynamics reveal a dissociation between storage and selective attention within working memory. *Scientific Reports*, *9*(1), 13499. https://doi.org/10.1038/s41598-019-49577-0

Hasson, U., Nastase, S. A., & Goldstein, A. (2020). Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron*, *105*(3), 416-434.

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., & Schultz, T. (2014). Mental workload during n-back task—Quantified in the prefrontal cortex using fNIRS. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00935

Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. *Human factors*, *46*(4), 611-624.

Hwang, G., Jacobs, J., Geller, A., Danker, J., Sekuler, R., & Kahana, M. J. (2005). EEG correlates of verbal and nonverbal working memory. *Behavioral and Brain Functions : BBF*, *1*, 20. https://doi.org/10.1186/1744-9081-1-20

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, *17*(3), 259–273. https://doi.org/10.1111/j.1469-8986.1980.tb00146.x

Jain, A. K., & Chandrasekaran, B. (1982). 39 Dimensionality and sample size considerations in pattern recognition practice. *Handbook of statistics*, *2*, 835-855.

Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kiroi, V. N., Aslanyan, E. V., Bakhtin, O. M., Minyaeva, N. R., & Lazurenko, D. M. (2016). EEG Correlates of the Functional State of Pilots during Simulated Flights. *Neuroscience and Behavioral Physiology*, *46*(4), 375–381. https://doi.org/10.1007/s11055-016-0245-6

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: Attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641). https://doi.org/10.1098/rstb.2013.0205

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, *133*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Liu, Y., Bengson, J., Huang, H., Mangun, G. R., & Ding, M. (2016). Top-down Modulation of Neural Activity in Anticipatory Visual Attention: Control Mechanisms Revealed by Simultaneous EEG-fMRI. *Cerebral Cortex (New York, NY)*, *26*(2), 517–529. https://doi.org/10.1093/cercor/bhu204

Liu, Y., & Wickens, C.D. (1992). Visual scanning with or without spatial uncertainty and divided and selective attention. Acta Psychologica, 79, 131-153.

Lotte, F. (2015). Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain–Computer Interfaces. *Proceedings of the IEEE*, *103*(6), 871–890. https://doi.org/10.1109/JPROC.2015.2404941

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, *4*(2), R1–R13. https://doi.org/10.1088/1741-2560/4/2/R01

Lotte, F, Larrue, F., & Mühl, C. (2013). Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: Lessons learned from instructional design. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00568

Mandrick, K., Chua, Z., Causse, M., Perrey, S., & Dehais, F. (2016). Why a Comprehensive Understanding of Mental Workload through the Measurement of Neurovascular Coupling Is a Key Issue for Neuroergonomics? *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00250

Mapelli, I., & Özkurt, T. E. (2019). Brain oscillatory correlates of visual short-term memory errors. *Frontiers in human neuroscience*, *13*, 33.

Martens, M., & Van Winsum, W. (2000). *Measuring distraction: The Peripheral Detection Task*.

Matthews, G., De Winter, J., & Hancock, P. A. (2020). What do subjective workload scales really measure? Operational and representational solutions to divergence of workload measures. *Theoretical Issues in Ergonomics Science*, 1-28.

Mcmahan, T., Parberry, I., & Parsons, T. D. (n.d.). *Evaluating Electroencephalography Engagement Indices During Video Game Play*.

Navon, D., & Gopher, D. (1979). On the economy of the human processing systems . Psychological Review, 86, 254-255.

Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. *sensors*, *12*(2), 1211-1279.

Nittala, S. K. R., Elkin, C. P., Kiker, J. M., Meyer, R., Curro, J., Reiter, A. K., Xu, K. S., & Devabhaktuni, V. K. (2018). Pilot Skill Level and Workload Prediction for Sliding-Scale Autonomy. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1166–1173. https://doi.org/10.1109/ICMLA.2018.00188

Oberauer, K. (2019). Working Memory and Attention – A Conceptual Analysis and Review. *Journal of Cognition*, *2*(1), 36. https://doi.org/10.5334/joc.58

Osalusi, B., Abraham, A., & Aborisade, D. (2018). EEG Classification in Brain Computer Interface (BCI): A Pragmatic Appraisal. *American Journal of Biomedical Engineering*, *8*(1), 1–11.

Palva, S., Kulashekhar, S., Hämäläinen, M., & Palva, J. M. (2011). Localization of Cortical Phase and Amplitude Dynamics during Visual Working Memory Encoding and Retention. *The Journal of Neuroscience*, *31*(13), 5013–5025. https://doi.org/10.1523/JNEUROSCI.5592-10.2011

Paus, T., Zatorre, R. J., Hofle, N., Caramanos, Z., Gotman, J., Petrides, M., & Evans, A. C. (1997). Time-related changes in neural systems underlying attention and arousal during the performance

of an auditory vigilance task. *Journal of Cognitive Neuroscience*, *9*(3), 392–408. https://doi.org/10.1162/jocn.1997.9.3.392

Pion-Tonachini, L., Hsu, S. H., Makeig, S., Jung, T. P., & Cauwenberghs, G. (2015, August). Real-time eeg source-mapping toolbox (rest): Online ica and source localization. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4114-4117). IEEE.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, *40*(1), 187–195. https://doi.org/10.1016/0301-0511(95)05116-3

Pulvermüller, F., Garagnani, M., & Wennekers, T. (2014). Thinking in circuits: Toward neurobiological explanation in cognitive neuroscience. *Biological Cybernetics*, *108*(5), 573–593. https://doi.org/10.1007/s00422-014-0603-9

Puma, S., Matton, N., Paubel, P.-V., Raufaste, É., & El-Yagoubi, R. (2018). Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology*, *123*, 111–120. https://doi.org/10.1016/j.ijpsycho.2017.10.004

Radüntz, T. (2017). Dual Frequency Head Maps: A New Method for Indexing Mental Workload Continuously during Execution of Cognitive Tasks. *Frontiers in Physiology*, *8*. https://doi.org/10.3389/fphys.2017.01019

Radüntz, T. (2020). The Effect of Planning, Strategy Learning, and Working Memory Capacity on Mental Workload. *Scientific Reports*, *10*(1), 7096. https://doi.org/10.1038/s41598-020-63897-6

Raudy, SJ., & Jain, AK. (1991). Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 3, pp. 252-264.

Rojas, R., Debie, E., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., Garratt, M., & Abbass, H. (2020). Electroencephalographic Workload Indicators During Teleoperation of an Unmanned Aerial Vehicle Shepherding a Swarm of Unmanned Ground Vehicles in Contested Environments. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00040

Saravini, F. (1999). Energy and the brain: Facts and fantasies. In E. Della Salla (Ed.), Mind myths (pp. 43–58). Chichester, England: Wiley.

Setareh, H., Deger, M., & Gerstner, W. (2018). Excitable neuronal assemblies with adaptation as a building block of brain circuits for velocity-controlled signal propagation. *PLOS Computational Biology*, *14*(7), e1006216. https://doi.org/10.1371/journal.pcbi.1006216

Song, J., Zhuang, D., Song, G., & Wanyan, X. (2011). Pilot mental workload measurement and evaluation under dual task. *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, *2*, 809–812. https://doi.org/10.1109/BMEI.2011.6098376

Stone, D. B., Tamburro, G., Fiedler, P., Haueisen, J., & Comani, S. (2018). Automatic removal of physiological artifacts in EEG: the optimized fingerprint method for sports science applications. *Frontiers in human neuroscience*, *12*, 96.

Tallon-Baudry, C., Kreiter, A., & Bertrand, O. (1999). Sustained and transient oscillatory responses in the gamma and beta bands in a visual short-term memory task in humans. *Visual Neuroscience*, *16*(3), 449–459. https://doi.org/10.1017/s0952523899163065

Tatum, W. O., Dworetzky, B. A., & Schomer, D. L. (2011). Artifact and recording concepts in EEG. *Journal of clinical neurophysiology*, *28*(3), 252-263.

Tauscher, J.-P., Schottky, F. W., Grogorick, S., Bittner, P. M., Mustafa, M., & Magnor, M. (2019). Immersive EEG: Evaluating Electroencephalography in Virtual Reality. *2019 IEEE Conference*

on Virtual Reality and 3D User Interfaces (VR), 1794–1800. https://doi.org/10.1109/VR.2019.8797858

Tăuţan, A.-M., Serdijn, W., Mihajlović, V., Grundlehner, B., & Penders, J. (2013). Framework for evaluating EEG signal quality of dry electrode recordings. *2013 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 186–189. https://doi.org/10.1109/BioCAS.2013.6679670

Tomioka, R., Dornhege, G., Nolte, G., Blankertz, B., Aihara, K., & Müller, K. R. (2006). Spectrally weighted common spatial pattern algorithm for single trial EEG classification. *Dept. Math. Eng., Univ. Tokyo, Tokyo, Japan, Tech. Rep*, *40*.

Tsubomi, H., Fukuda, K., Watanabe, K., & Vogel, E. K. (2013). Neural Limits to Representing Objects Still within View. *Journal of Neuroscience*, *33*(19), 8257–8263. https://doi.org/10.1523/JNEUROSCI.5348-12.2013

United States Bureau of Transportation Statistics. (2019). *National Transportation Statistics (series)*. https://doi.org/10.21949/1503663

Van Benthem, K. D., Herdman, C. M., Tolton, R. G., & LeFevre, J.-A. (2015). Prospective Memory Failures in Aviation: Effects of Cue Salience, Workload, and Individual Differences. *Aerospace Medicine and Human Performance*, *86*(4), 366–373. https://doi.org/10.3357/AMHP.3428.2015

Vanneste, P., Raes, A., Morton, J., Bombeke, K., Van Acker, B. B., Larmuseau, C., Depaepe, F., & Van den Noortgate, W. (2020). Towards measuring cognitive load through multimodal physiological data. *Cognition, Technology & Work*. https://doi.org/10.1007/s10111-020-00641-0

Wahn, B., & König, P. (2017). Is Attentional Resource Allocation Across Sensory Modalities Task-Dependent? *Advances in Cognitive Psychology*, *13*(1), 83–96. https://doi.org/10.5709/acp-0209-2

Walter, C., Schmidt, S., Rosenstiel, W., Gerjets, P., & Bogdan, M. (2013, September). Using cross-task classification for classifying workload levels in complex learning tasks. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 876-881). IEEE.

Wang, Y.-K., Jung, T.-P., & Lin, C.-T. (2018). Theta and Alpha Oscillations in Attentional Interaction during Distracted Driving. *Frontiers in Behavioral Neuroscience*, *12*. https://doi.org/10.3389/fnbeh.2018.00003

Wanyan, X., Zhuang, D., & Zhang, H. (2014). Improving pilot mental workload evaluation with combined measures. *Bio-Medical Materials and Engineering*, *24*(6), 2283–2290. https://doi.org/10.3233/BME-141041

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159–177.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), Varieties of attention (pp. 63-101). New York: Academic Press.

Wickens, C.D. (2008). Multiple resources and mental workload. Human Factors, 50, 449-455.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science*, *11*(4), 128–133. https://doi.org/10.1111/1467-8721.00184

Wickens, C. D., Harwood, K., Segal, L., Tkalcevic, I., & Sherman, B. (1988). TASKILLAN: A Simulation to Predict the Validity of Multiple Resource Models of Aviation Workload. *Proceedings of the Human Factors Society Annual Meeting*, *32*(2), 168–172. https://doi.org/10.1177/154193128803200237

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering Psychology and Human Performance*. Psychology Press.

Wickens, C.D. , & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. Human Factors, 30, 599-616.

Wrobel, A. (2000). Beta activity: A carrier for visual attention. *Acta Neurobiologiae Experimentalis*, *60*, 247–260.

Yamada, T., & Meng, E. (2012). *Practical guide for clinical neurophysiologic testing: EEG*. Lippincott Williams & Wilkins.

Zarjam, P., Epps, J., & Chen, F. (2011). Characterizing working memory load using EEG delta activity. *European Signal Processing Conference*.

Appendix

Appendix A: Additional analyses

Figure 17 shows the mean effect size of workload on EEG power density at each electrode site. Figure 18 shows participant effect sizes of workload on frontal theta power, which is another index of workload.

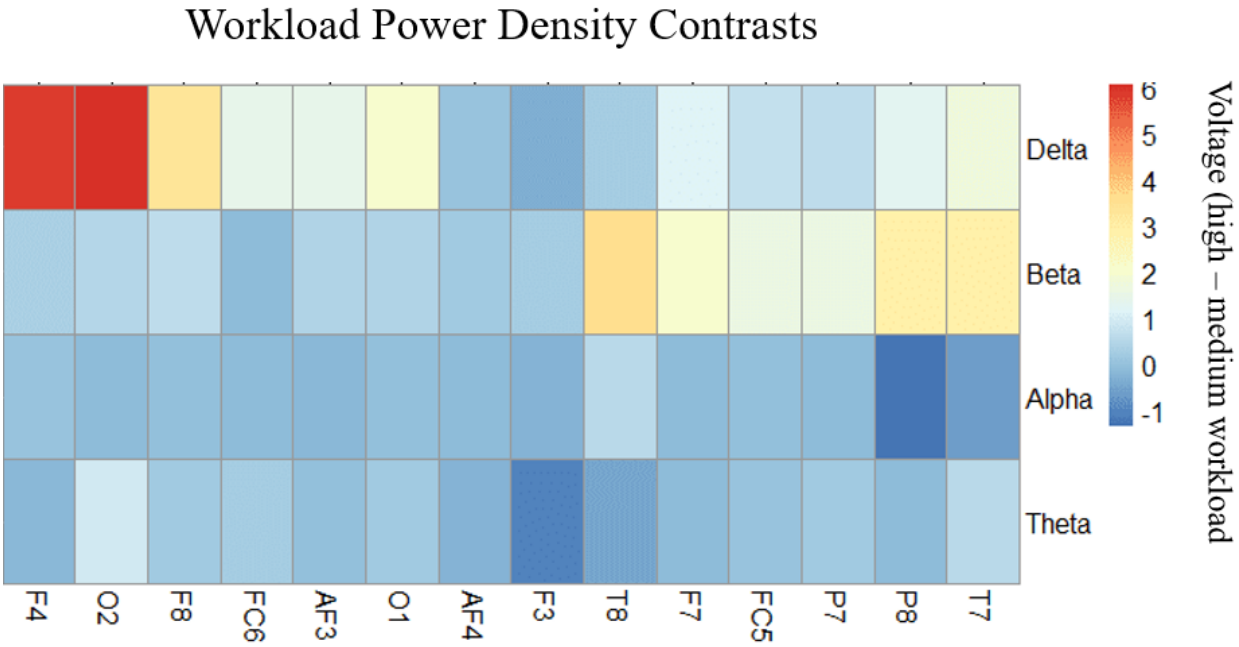

Figure 17. Displays the effect of workload on power spectral densities at each frequency band and each electrode. Effects sizes were measured by subtracting the absolute voltage in MWL from HWL conditions.

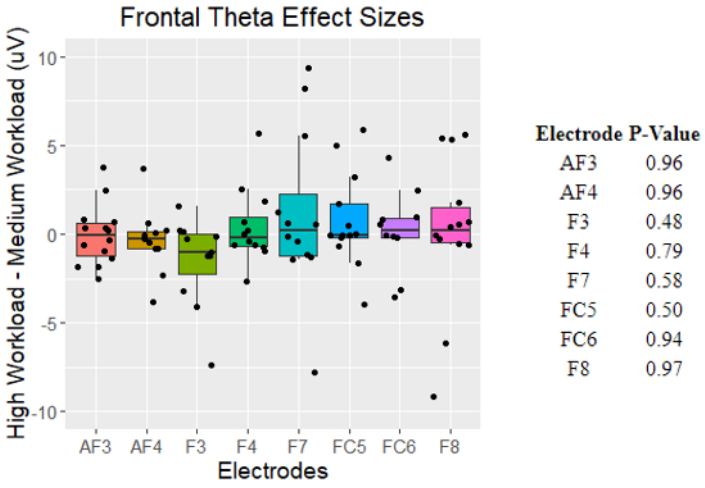

Figure 18. Distribution of participant workload power densities contrasts at frontal electrodes within the Theta ranges. Elevated frontal Theta power densities are a proposed index of workload (Radunzt, 2020).

Appendix B: Analysis Scripts

The automated preprocessing script can be downloaded at https://akfraser96.github.io. The electrode locations file will be required for running this script, which can also be obtained from this site.