Unravelling the Physiological Correlates of Mental Workload Variations in Tracking and Collision Prediction Tasks: Implications for Air Traffic Controllers

Alka Rachel John[1*], Avinash K Singh[1], Tien-Thong Nguyen Do[1], Ami Eidels[2], Eugene Nalivaiko[3], Alireza Mazloumi Gavgani[3], Scott Brown[2], Murray Bennett[2], Sara Lal[4], Ann M. Simpson[4], Sylvia M Gustin[5,6], Kay Double[7], Frederick Rohan Walker[3], Sabina Kleitman[8], John Morley[9], Chin-Teng Lin[1]

[1] Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney

[2] School of Psychology, University of Newcastle

[3] School of Biomedical Sciences and Pharmacy and Centre for Advanced Training Systems, University of Newcastle

[4] School of Life Sciences, University of Technology Sydney

[5] School of Psychology, University of New South Wales

[6] Centre for Pain IMPACT, Neuroscience Research Australia

[7] Brain and Mind Centre and Discipline of Pharmacology, School of Medical Sciences, The University of Sydney

[8] Faculty of Science, The University of Sydney

[9] School of Medicine, Western Sydney University

*Corresponding Author: Alka Rachel John, alkarachel.john@student.uts.edu.au

Abstract

The surge in air traffic increases the workload experienced by air traffic controllers (ATC) while they organise traffic-flow and prevent conflicts between aircraft. Even though several factors influence the complexity of ATC tasks, keeping track of the aircraft and preventing collision are the most crucial. We have designed tracking and collision prediction tasks to elucidate the differences in the physiological response to the workload variations in these basic ATC tasks to untangle the impact of workload variations experienced by operators working in a complex ATC environment. Physiological measures, such as electroencephalogram (EEG), eye activity, and heart rate variability (HRV) data, were recorded from 24 participants performing tracking and collision prediction tasks with three levels of difficulty. The mental workload in the tracking task was found to be positively correlated with the frontal theta power and negatively correlated with the occipital alpha power. In contrast, for the collision prediction task, the frontal theta, parietal theta, occipital delta, and theta power were positively correlated, and parietal alpha power was negatively correlated with the increases in mental workload. The pupil size, number of blinks and HRV metric, root mean square of successive difference (RMSSD), also varied significantly with the mental workload in both these tasks in a similar manner. Our findings indicate that variations in task load are sensitively reflected in physiological signals, such as EEG, eye activity and HRV, in these basic ATC-related tasks. Furthermore, the markedly distinct neurometrics of workload variations in the tracking and collision prediction tasks indicate that neurometrics can provide insights on the type of mental workload. These findings have applicability to the design of future mental workload adaptive systems that integrate neurometrics in deciding not just 'when' but also 'what' to adapt. Our study provides compelling evidence in the viability of developing intelligent closed-loop mental workload adaptive systems that ensure efficiency and safety in ATC and beyond.

Keywords: Air traffic controller, Mental workload, EEG, pupil size, blink rate, RMSSD

Introduction

People tend to avoid performing tasks that push their capabilities beyond their limits as they find it frustrating and stressful (Ahlstrom, 2010). However, not all work environments offer that luxury, which makes it crucial to establish good interaction between the human operator abilities and work environment (Wickens et al., 2015). Even though human operators can easily adapt to diverse work environments and perform several tasks and use different equipment simultaneously, poorly designed work environments cause an overload of sensory information resulting in excess workload. Air traffic controllers operate in such a complex environment to ensure a safe and efficient air traffic flow by organising traffic flow in a way that aircraft reach their destination in a well-organized and expeditious manner. It is their job to anticipate and prevent conflicts between aircraft by monitoring whether aircraft adhere to the International Civil Aviation mandated separation standards (Rodgers and Drechsler, 1993), and managing the resulting complexity. They routinely manage a significant number of aircraft, coming from different directions and heading to various destinations at diverse speeds and altitudes (Gronlund et al., 1998). However, as the air traffic increases, there is a growing need to study the mental factors that ensure the efficiency of air traffic controllers.

Mental workload is one of the most crucial factors that affects the efficiency of air traffic controllers as they operate in complex interactive work environments. In recent years ATCs roles have transitioned to a supervisory level, and they must now integrate multiple streams of information. This demands more cognitive resources (Pashler, 1994), resulting in a higher workload for operators (Kompier et al., 2001; Niosh, 2002; Landsbergis, 2003). Wickens and Tsang, 2015 defined *mental workload* as the dynamic relationship between the cognitive resources demanded by a task and the capability of the operator to afford those resources. Human operators have limited information processing abilities as they have finite resources with limited capacity (Kahneman, 1973; Kramer and Spinks, 1991).

The theory of limited cognitive resources states that exposure to demanding task conditions impairs performance due to resource depletion (PA Van Dongen et al., 2011) or compromised access to resources (Borragan Pedraz, 2016). As mental workload has a negative influence on the performance of the operator, it results in human error commission (Reason, 2000), compromising system efficiency and safety (Xie and Salvendy, 2000). The mental workload of the operator should be at optimal level avoiding both underload and overload (Hancock, 1989, Borghini et al., 2014) as the performance of the operator is known to fall at both overload and underload conditions (Yerkes and Dodson, 1908, Calabrese, 2008, Van Acker et al., 2018, Hancock and Matthews, 2019). As the dynamic adaptive theory states, the brain seeks resource homeostasis and cognitive comfort and extremely high and low task demands degrades adaptability and thereby, performance (Hancock and Warm, 2003). Predicting the mental workload of the operator, and thereby adapting system behaviour by modifying task allocation can avoid this loss of situational awareness, maintaining high performance. Accurate and reliable measurement of mental workload of the operator is crucial especially in a safety-critical work environment by providing better work environments and human-machine interactions (Byrne and Parasuraman, 1996, Scerbo, 2001, Aricò et al., 2017).

Researchers have relied on multiple strategies, such as self-assessment, performance measures and physiological metrics, to assess mental workload; however, each of these methods has their benefits and drawbacks (O'Donnell, 1986, Wierwille and Eggemeier, 1993). The assessment from these different mental workload measurement methods is often dissociated (Yeh and Wickens, 1988) as the sensitivity of these measures are heavily dependent on the operator's workload (De Waard and Groningen, 1996). Several subjective measures such as the Instantaneous self-assessment (ISA) questionnaire (Brennan, 1992, Jordan, 1992, Kirwan et al., 1997), NASA Task Load Index (Hart and Staveland, 1988), the Subject Workload Assessment Technique (Reid et al., 1988), are used to access the workload of the operator. Mental workload is a complex construct, which reflects the available cognitive resources and cannot be accurately assessed using subjective measurements alone (de Waard and Lewis-Evans, 2014). Moreover, the mental workload assessment method should not disrupt the task at hand or influence the mental state of the operator, which might be the case of subjective assessment strategies using questionnaires. Another widely used workload assessment method is the performance-based workload measurement, which will, like the subjective assessment method, provide a retrospective workload assessment. Moreover, performance-based measures can only tell part of the story as operators can achieve the same performance experiencing higher workload (Aricò et al., 2019). Over the years, physiological metrics have been used to assess workload (Casali and Wierwille, 1984, Matthews et al., 2015, Charles and Nixon, 2019) as it offers high sensitivity,

diagnostic ability and is mostly non-intrusive (Parasuraman and Rizzo, 2008, Zhao et al., 2018), providing an accurate and real-time assessment of the operator's workload. The use of physiological data such as neurophysiological signals can assess mental workload online without influencing the task as there is no explicit output (Parasuraman, 2001, Gevins and Smith, 2003, Parasuraman, 2015). Neurophysiological measures can also assess the changes in the mental state that are not merely discernible in overt task performance (Parasuraman, 2015, Wickens and Tsang, 2015, Aricò et al., 2016, Blankertz et al., 2016).

Neurophysiological measures, such as the electroencephalogram (EEG) signal has been widely employed to estimate mental workload as the effects of task demand are clearly visible in EEG rhythm variations (Brookings et al., 1996, Gevins and Michael Smith, 2003, Borghini et al., 2014, Borghini et al., 2014, Di Flumeri et al., 2014, Lopez-Gordo et al., 2014, Matthews et al., 2015, Radüntz, 2018, Radüntz and Meffert, 2019; Lin and Do, 2021). Researchers have also used EEG to reliably predict performance degradation from workload variations (Matousek and Petersén, 1983, Gevins et al., 1990) and noted that it is correlated with an increase in the frontal theta power and a change in parietal alpha power, which relates to cognitive and memory performance (Gale et al., 1977, Sterman and Mann, 1995, Pfurtscheller, 1997, Gevins et al., 1998, Gevins and Smith, 2000). Many EEG-based workload indices, such as the ratio of frontal theta to parietal alpha power (Fritz et al., 2014, Holm et al., 2009), the ratio of beta to theta and alpha (Freeman et al., 1999), theta-beta ratio (Montgomery et al., 1998) reliably reflect workload. However, EEG features of the mental workload are found to be task-dependent, therefore, adding other modalities like eye activity data and heart rate data can help achieve far superior outcomes (Ke et al., 2014, Popovic et al., 2015). Pupil size and blink rate have recently attracted attention as a reliable indicator of workload (Marquart et al., 2015). Heart rate variability (HRV) is yet another highly sensitive physiological index to mental workload variations (Kamath and Fallen, 1993, Nickel and Nachreiner, 2003, Hjortskov et al., 2004, Murai et al., 2004, Murphy et al., 2004). Root mean square of successive differences (RMSSD) is the most robust time-domain HRV measure of workload (Mehler et al., 2011).

Once the mental workload of the operator can be reliably assessed, it can be used to drive a mental workload adaptive system (Prinzel et al., 2000; Schmorrowe et al., 2006). In such adaptive systems, the physiological measures of mental workload can be used to trigger the adaptive automation that will adapt its behaviour to the current mental workload of the operator (Scerbo, 1996; Kaber and Endsley, 2004). A mental workload adaptive automation system should be able to conform to the variations in the mental workload of the operator without them having to explicitly state their needs or triggering the automation. When human operators and automation team up to achieve better performance and efficiency, the operator expects automation to behave like a human coworker (Aricò et al., 2017). Therefore, adaptive automation should be timely, stepping in at the right time and cognitively empathetic with the operator, helping where it is needed, taking over the task that is currently overwhelming the operator. However, currently, physiological correlates of the mental workload are only used to decide 'when' to adapt and not 'what' to adapt, keeping the strategies employed by the adaptive automation system still primitive. There is a need to develop intelligent adaptive systems that can identify what form of automation to use depending on the type of mental workload experienced by the operator. Nonetheless, there is still a dearth in evidence that physiological metrics of mental workload can direct to the tasks contributing to workload.

In this paper we investigated whether the multimodal physiological metrics of mental workload can provide more information about the task contributing to the workload experienced by the ATC operator. Even though several factors influence the complexity of ATC tasks (Mogford et al., 1995, Cummings and Tsonis, 2005), such as environmental, display, traffic and organisational factors, the main functions for ATC operator are tracking and collision prediction. Therefore, we designed tracking and collision prediction tasks to elucidate the physiological effects of workload variations in these basic ATC tasks. The experiment was fashioned as a classical cognitive paradigm with a manipulation of workload (low, medium, high) and repeated stimuli to study whether physiological data such as EEG, eye activity and HRV can reliably assess the mental workload of the operator while they perform these basic tracking and collision prediction tasks. We formulated the following four research hypotheses for our study:

H1. The three distinct levels of workload defined in both tracking and collision prediction tasks can yield significant performance degradation with the increasing levels of workload.

H2. Workload variation in tracking and collision prediction tasks can be reliably assessed using EEG, eye activity and HRV metrics.

H3. The performance in tracking and collision prediction tasks can be predicted based on the measured physiological signals.

H4. Physiological response to the workload variations in the tracking and collision prediction tasks will be distinct across tasks.

Methods

Participants

Twenty-four participants (age 25 ± 5, 17 males and 7 females, all right-handed) participated in this experiment after giving written informed consent. All the participants had a normal or corrected vision and no history of any psychological disorder that might affect the results. The experimental protocol was approved by the University of Technology Sydney Human Research Ethics Expedited Review Committee (ETH19-4197).

The EEG data were collected using SynAmps2 Express system (Compumedics Ltd., VIC, Australia) with 64 Ag/AgCl sensors system. The placement of these electrodes was consistent with the extended 10% system (Chatrian and Nelson, 1985) and the impedance of each electrode was ensured to be below 10 kΩ before each session. The data were collected at a sampling rate of 1000 Hz. Eye activity data was collected using Pupil Labs Pupil Core (Berlin, Germany). This wearable eye-tracking headset has three cameras, two of which record the eyes activity at 200 Hz sampling rate, and the other one records the participant's field of view at 30 Hz sampling rate (Kassner et al., 2014). The Blood Volume Pulse (BVP) data was recorded using infrared plethysmography-based Empatica E4 (Empatica Srl, Milano, Italy). The real-time synchronisation of events from the task scenario to the EEG, eye activity and BVP data was achieved by the Lab Streaming Layer (Kothe, 2015).

Experimental Procedures

Our experimental design included two tasks – multiple objects tracking task (Innes et al., 2019) and collision prediction task. As shown in Figure 1(A), in the tracking task, during the initial 3 seconds, participants look at a fixation cross on the screen followed by a freeze phase, where the dots, some of which are blue, and the rest are red, remain stationary. The blue dots are the dots that need to be tracked (hence, 'targets'). After three seconds of freeze, the blue targets also turn red so that they are no longer distinctive from the other dots and all the dots start moving. Each dot has a diameter of 14 pixels, and they move randomly in the display area at a frame rate of 15 frames/second. The participant is asked to keep track of the targets (dots that were initially blue) for 15 seconds. After this time window all dots stop moving and the participants should indicate the target dots by clicking on the dots that they have kept track of. The workload levels in this tracking task are manipulated by varying the number of blue dots and the total number of dots. As shown in Table 1, for the low workload condition, there are 10 dots with one blue dot. In medium workload, there are 12 dots with three blue dots, and for high workload, there are 15 dots with five blue dots.

As shown in Figure 1(B), in the collision prediction task, there is a fixation cross on the screen for three seconds. Then there is a three-second-long freeze phase where the dots remain stationary, after which all the dots start moving. Unlike the tracking tasks, all dots are of the same color (pink). The participant is required to predict the trajectory of the dots and identify which pair of dots would collide. Dots move in a predefined uniform direction, and we have manipulated the trajectory of the dots such that there will be only one collision in each trial. The participants were asked to identify the pair of dots that would collide and click on both dots before the collision happens, which usually occurs in the last 3 seconds of the trial. In order to prevent random guesses, the number of dots the participants can select is limited to two, and once the participant clicks on the dot, it changes from pink to red colour. The levels of workload were manipulated by varying the number of dots. The low workload condition had six dots while the medium workload had 12 dots and 18 dots for high workload conditions, as shown in Table 1. A 15-inch monitor with 1920 x 1080 resolution was used to display both these tasks.

Each participant had to perform 108 trials of each task with 36 trials of each workload level. The entire experiment was divided into four blocks and each block had 27 trials of the tracking task and 27 trials of the collision prediction task. The type of workload condition in the trials was randomised within a block to avoid any habituation or expectation effects. At the end of each trial, the participant received feedback on their performance with the following message on the screen– "You have correctly tracked x dots out of y dots to track" for the tracking task and "You have correctly detected this collision" or "You have missed this collision" for the collision prediction task. They could move to the next trial after reading the performance feedback by pressing the spacebar key. After each block, the participants were advised to rest for 5 minutes before proceeding to the next block by pressing the spacebar key. All participants were trained in a training session which usually lasted approximately ten minutes, where they performed six trials of each task to familiarise themselves with the tasks and develop strategies for successfully executing the tasks. The participants were asked to continue the training until they felt comfortable with the tasks. After the training, all participants performed the tasks for ~ 1.5 hours during which EEG, eye activity and HRV data were collected.

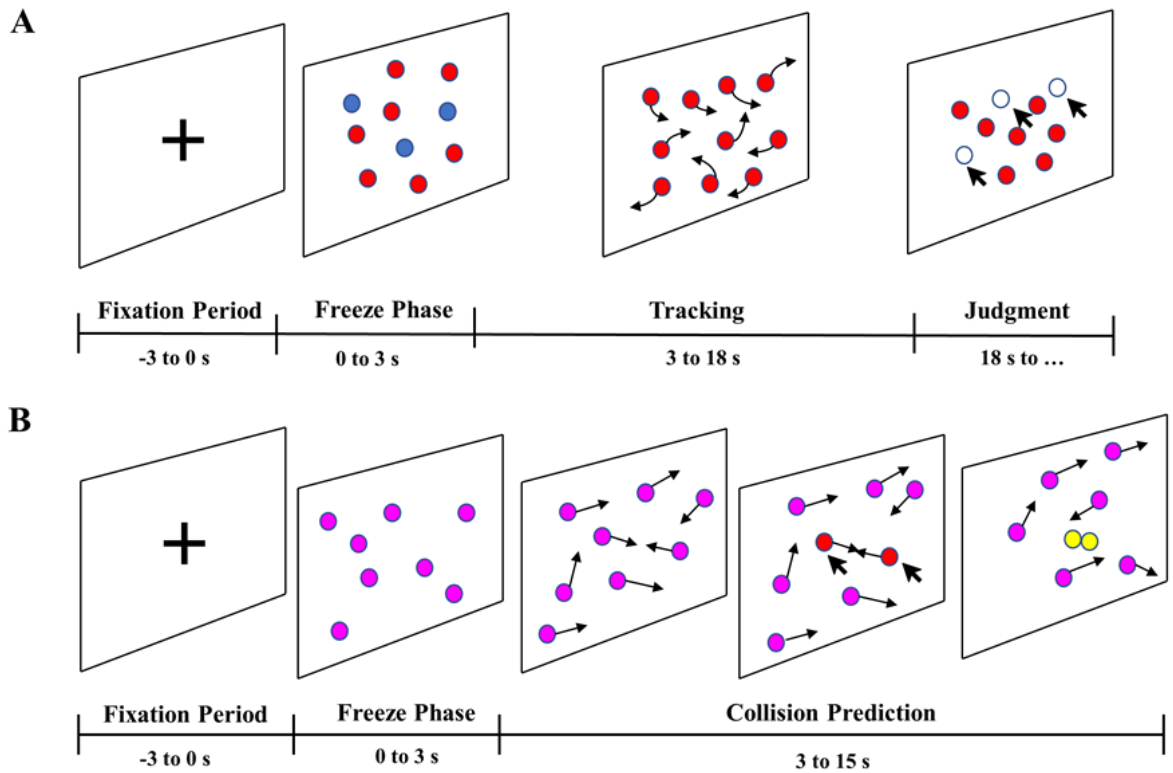


Figure 1: The experimental design of the tasks. (A) the experimental design of the tracking task and (B) shows the design of the collision prediction task. The number of dots shown in these diagrams is just for representation purposes.

Table 1: Workload Manipulations in the tracking and collision prediction tasks

TASK	WORKLOAD LEVEL	WORKLOAD MANIPULATION	
		Tracking Dots	Total Number of Dots
Tracking Task	Low	1	10
	Medium	3	12
	High	5	15
		WORKLOAD MANIPULATION	
		Total Number of Dots	
Collision Prediction Task	Low	6	
	Medium	12	
	High	18	

Data Analysis

Behavioural and Performance Data Analysis

For the tracking task, each participant's performance was evaluated by examining the tracking accuracy. Tracking accuracy for each trial was defined as the ratio of the number of correctly tracked dots to the total number of dots to track.

$$Tracking\ Accuracy = \frac{Number\ of\ Correctly\ Tracked\ Dots}{Total\ Number\ of\ Dots\ to\ Track} \quad (1)$$

In case of the collision prediction trials, the performance was determined using the time before collision and collision miss proportion rate. The time before collision is the time period between when the participant clicks on either one of the colliding dots and when the collision happens (see Supplementary Figure 1). The collision miss proportion rate for a particular workload level of the collision prediction task is the ratio of the number of collision prediction misses to the total number of collisions in that specific workload level. A collision miss was considered to happen when the participant was unable to identify which pair of dots would collide and, hence, did not click on either of the dots before the collision.

$$Collision\ Prediction\ Miss\ Proportion\ Rate = \frac{Number\ of\ Missed\ Collision\ Predictions}{Total\ Number\ of\ Collisions} \quad (2)$$

EEG Preprocessing

EEG data were preprocessed using EEGLAB v2020.0 toolbox (Delorme and Makeig, 2004) in MATLAB R2019a (The Mathworks, Inc., Natick, MA, USA), and adapted from Do et al., 2020 (see Supplementary Figure 2). EEG data were down-sampled to 250 Hz, and a band-pass filter of 2–45 Hz was applied. Channels with three seconds or more flat line were removed using the clean_flatline function. Noisy channels were identified and removed using the clean_channels function in EEGLAB. On an average 3±1 channels were removed and these channels were restored by interpolating the data from neighbouring channels using the spherical spline method from the EEGLAB toolbox. Continuous artifactual regions were removed using the EEGLAB function, pop_rejcont. The data was divided into epochs of length 0.5 seconds, with an overlap of 0.25 seconds and the frequency threshold considered was 1 to 100 Hz. Each selected artifactual region consisted of at least four contiguous epochs with high-frequency data (spectrum over 10 dB). Then window cleaning was performed using the clean_windows function in EEGLAB. This function computes the power of each sliding window of one second length and transforms it to z-score to reject all windows in which the computed value lies outside 5 standard deviations. After these artifact removal steps, two EEG datasets were extracted, one comprising tracking trials and one with the collision prediction trials. Each participant had 34±2 high workload, 35±1 medium workload and 34±1 low workload tracking trials, and 32±2 high workload, 33±2 medium workload and 33±1 low workload collision prediction trials.

The tracking epochs were 21 seconds long and included the 3 seconds of the fixation period followed by 3 seconds of freeze, after which the tracking task was commenced. The collision prediction task epochs were of 15 seconds length and included the initial 3 seconds of fixation period followed by 3 seconds of freeze phase and then, the collision prediction task. Both tracking and collision prediction datasets were decomposed using Independent Component Analysis (ICA), performed using EEGLAB's runica algorithm (Delorme and Makeig, 2004). Finally, we employed ICLabel (Pion-Tonachini et al., 2019), an automatic IC classifier to identify components related to brain, heart, line noise, eye, muscle, channel noise and other activities. This tool was adopted to generate class labels for each component and all the components with labels other than brain activity were rejected.

IC Clustering

EEGLAB STUDY structure (Delorme et al., 2011) was used to manage and process data recorded from multiple participants as it provides component clustering to cluster similar independent components across participants and allows statistical comparisons of component activities for different workload conditions. Clustering functions were used to examine the contributions of frontal and parietal clusters of independent components (ICs) to the workload dynamics. Frontal and parietal brain regions have been reported to reflect the changes in workload (Brookings et al., 1996; Shou et al., 2012; Matthews et al., 2015; Aricò et al., 2016; Aricò et al., 2017; Radüntz et al., 2017; Di Flumeri et al., 2018), and as both our tasks also manipulate the visual load, we particularly focused on the frontal, parietal and occipital clusters of brain activity.

A Study was created for each task, and each Study had one group (with 24 participants) with three conditions corresponding to the three levels of workload. Since the dataset of each participant was recorded in a single session, the resulting independent component maps were the same across all the three conditions for each participant. For each participant, only those ICs that had a residual variance (RV) less than 15% and inside the brain volume were chosen, which was achieved using Fieldtrip extension (Oostenveld et al., 2011). The k-means clustering algorithm (Hartigan and Wong, 1979) was used to cluster independent components across all participants to clusters based on two equally weighted (weight=1) criteria: (1) scalp maps and (2) their equivalent dipole model locations, which was performed using DIPFIT routines (Oostenveld and Oostendorp, 2004) in EEGLAB. Talairach coordinates (Lancaster et al., 2000) of the fitted dipole sources of these clusters were identified to select frontal, parietal and occipital clusters.

The grand-mean IC event-related spectral power changes (ERSPs) for each condition was subsequently calculated for each cluster. ERSPs (Makeig, 1993) shows the relative change in power at components with respect to a baseline period before the stimulus. The three seconds of fixation phase in each tracking and collision prediction epoch was taken as the baseline to see the changes in power spectra during the task. ERSPs for frontal, parietal and occipital clusters for both tracking and prediction tasks were examined. To compare the ERSP of different workload conditions, permutation-based statistics, implemented in EEGLAB, was used with Bonferroni correction and significance level set to $p = .05$. Also, for the frontal, parietal and occipital cluster, each ICs' spectral powers were calculated using EEGLAB’s spectopo function, which uses Welch’s periodogram method (Welch, 1967) on each 2-s segment using a Hamming window with 25% overlap for a range of frequencies from 2 to 45 Hz. For each IC, the power spectral density (PSD) at different frequency bands were examined to identify the correlates of mental workload.

Eye Activity data

Pupil Core software, Pupil Capture provides the pupil size for the left and right eye separately along with the associated confidence value, which represents the quality of the detection result. All data points where the confidence of the pupil size was less than 0.8 were removed from the data. The pupil size data was low pass filtered (using a minimum order finite impulse response filter) at 4 Hz (Privitera et al., 2010). As pupil size is a continuous measurement which is idiosyncratic and varies across participants, the raw pupil size data was normalised using the baseline data (defined as the three seconds of fixation period in each tracking and collision

prediction epoch). The number of blinks during each trial was also extracted from the pupil size measurement when the pupil size and confidence of the measurement, reported by the Pupil Capture software, suddenly dropped to zero.

Heart Rate Variability

Inter-beat-interval (IBI) time series was computed from the Blood Volume Pulse (BVP) data of each tracking and collision prediction trial. Root Mean Square of the Successive Differences (RMSSD) was computed by detecting peaks of the BVP using PeakUtils Python package (Negri, 2018) and calculating the lengths of the intervals between adjacent beats.

$$RMSSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(IBI_{i-1} - IBI_i)^2} \quad (3)$$

RMSSD data was also normalised by considering the three seconds of fixation period in each tracking and collision prediction epoch as the baseline.

Statistical Analysis

Statistical analyses were carried out using the SPSS (IBM SPSS 26.0; Chicago, IL, U.S.A.) statistical tool. In order to investigate the differences in the performance, EEG, eye activity and HRV parameters across participants in the three workload levels of tracking and collision prediction tasks, one-way repeated-measures analysis of variance (ANOVA) was conducted with workload level (low, medium or high) as the within-subjects factor. Mauchly's test was implemented to test for sphericity. We performed Greenhouse-Geisser correction if sphericity was not satisfied ($p < .05$). If the main effect of the ANOVA was significant, post-hoc comparisons were made to determine the significance of pairwise comparisons, using Bonferroni correction. Finally, multiple linear regression was performed to relate EEG, eye activity and HRV metrics to the performance in the tracking and collision prediction tasks. EEG power, eye activity and HRV metrics were all entered as predictors using the enter method, and the performance in the task was the dependent variable.

Results

Behavioural and Performance Measures

In the tracking task, tracking accuracy decreased significantly with increasing levels of workload, as shown in Figure 2(A). A repeated-measures ANOVA showed that tracking accuracy differed significantly between workload conditions [$F(2, 54) = 239.910$, $p < .001$, $\eta_p^2 = .899$]. Tracking accuracy during low workload was significantly higher than medium ($p < .001$) and high workload condition ($p < .001$). The tracking accuracy during medium workload was considerably higher than the high workload condition ($p < .001$).

For the collision prediction task, the time before collision and collision prediction miss proportion rate was considered. The time before collision decreased with increasing workload, as shown in Figure 2(B1). A repeated-measures ANOVA was conducted to study the effect of workload variations on time before collision, and the results showed that time before collision

varied significantly between workload conditions [F(1.497, 40.406) = 132.688, $p < .001$, $\eta_p^2 = .831$]. The time before collision decreased considerably in the medium ($p < .001$) and high ($p < .001$) workload conditions as compared to low workload condition. The time before collision during medium workload was also significantly greater than that during high workload condition ($p = 0.001$).

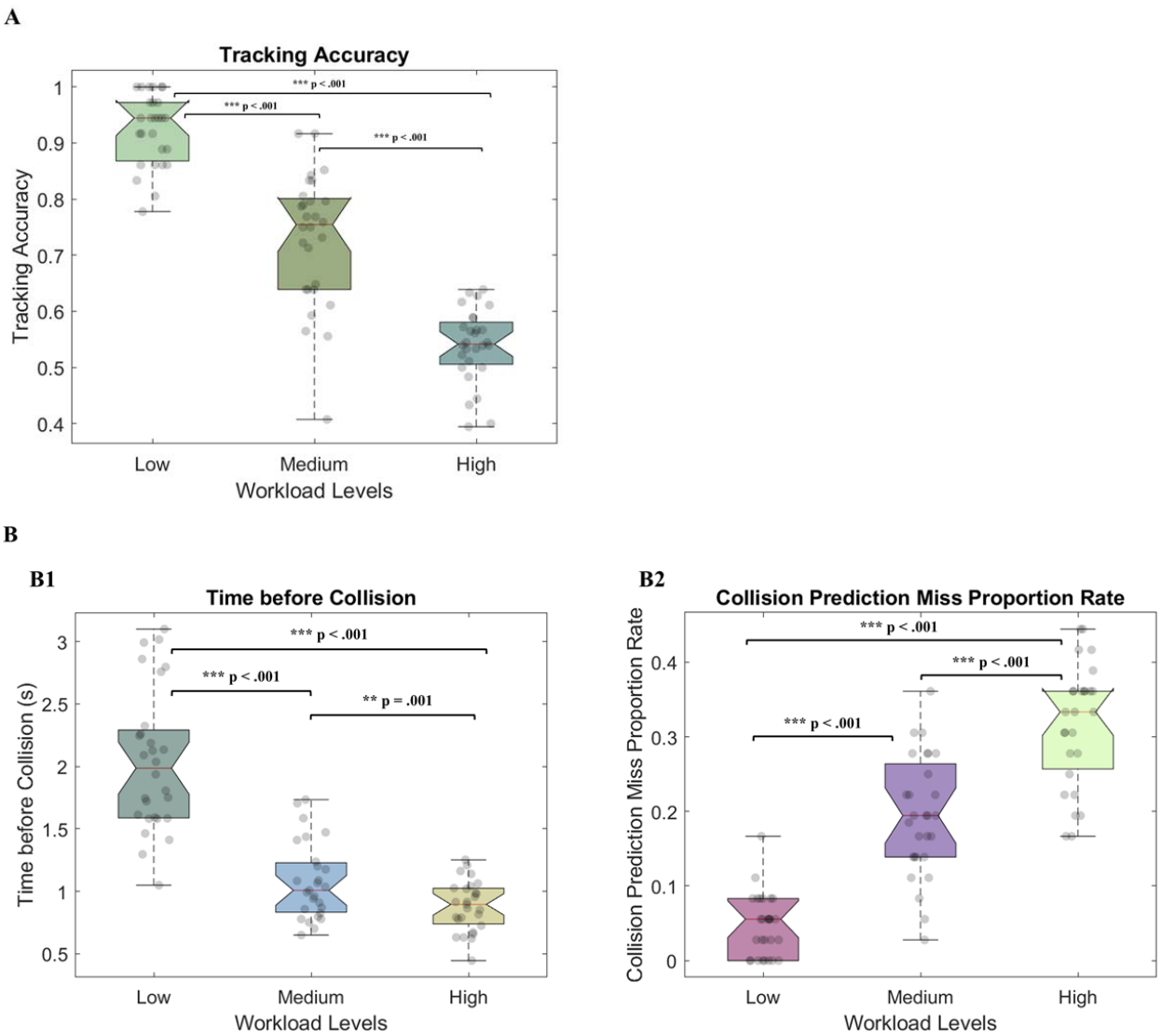


Figure 2: (A) shows the tracking accuracy of all the participants in the tracking task for the three levels of workload. (B) shows the performance of all participants in the collision prediction task for the three levels of workload. (B1) shows the mean time before collision for all the participants in the low, medium, and high workload conditions. (B2) shows the collision prediction miss proportion rate for the three levels of workload.

The collision prediction miss proportion rate increased with increasing levels of workload, as shown in Figure 2(B2). One-way repeated-measures ANOVA showed that the collision prediction miss proportion varied significantly between workload conditions [F(1.593, 43.009) = 116.338, $p < .001$, $\eta_p^2 = .812$]. The prediction miss proportion rate was markedly higher in the

medium ($p < .001$) and high ($p < .001$) workload conditions as compared to low workload condition. The collision prediction miss proportion rate during high workload was also significantly greater than that during medium workload condition ($p < .001$).

EEG Results

Independent Source Clusters

The frontal, parietal and occipital clusters were selected based on the location of fitted dipole sources (Oostenveld and Oostendorp, 2004). For the tracking task (refer Figure 3), the Talairach coordinate of the frontal cluster centroid was at (-1, 41, 27), the Talairach coordinate of the parietal cluster centroid was at (4, -51, 39) and the Talairach coordinate of the occipital cluster centroid was at (30, -70, 15). For the collision prediction task (see Figure 4), the Talairach coordinate of the frontal cluster centroid was at (-10, 17, 46), the Talairach coordinate of the parietal cluster centroid was at (5, -47, 47) and the Talairach coordinate of occipital cluster centroid was at (-3, -69, 20).

ERSP Changes with Mental Workload

Figures 5 illustrates frontal, parietal and occipital clusters' ERSP changes for three workload conditions: low, medium and high during the tracking task. Statistical analysis on ERSP changes of the frontal cluster (Figure 5(A1)) revealed a significant increase in theta power from the low level to the high level ($p < .05$). Figure 5(A2) shows the significant increase in theta power at the frontal cluster during high workload condition as compared to the medium workload condition. The frontal theta power was significantly greater than the low workload condition, as shown in Figure 5(A3).

However, no significant spectral power variations were observed at the parietal cluster. Figure 5(B1) shows the ERSP changes at the occipital cluster, which revealed a significant decrease in alpha power from the low level to the high level ($p < .05$). Figure 5(B2) shows the significant decrease in alpha power at the occipital cluster during high workload condition as compared to the medium workload condition. The occipital alpha power was significantly lesser than the low workload condition, as shown in Figure 5(B3).

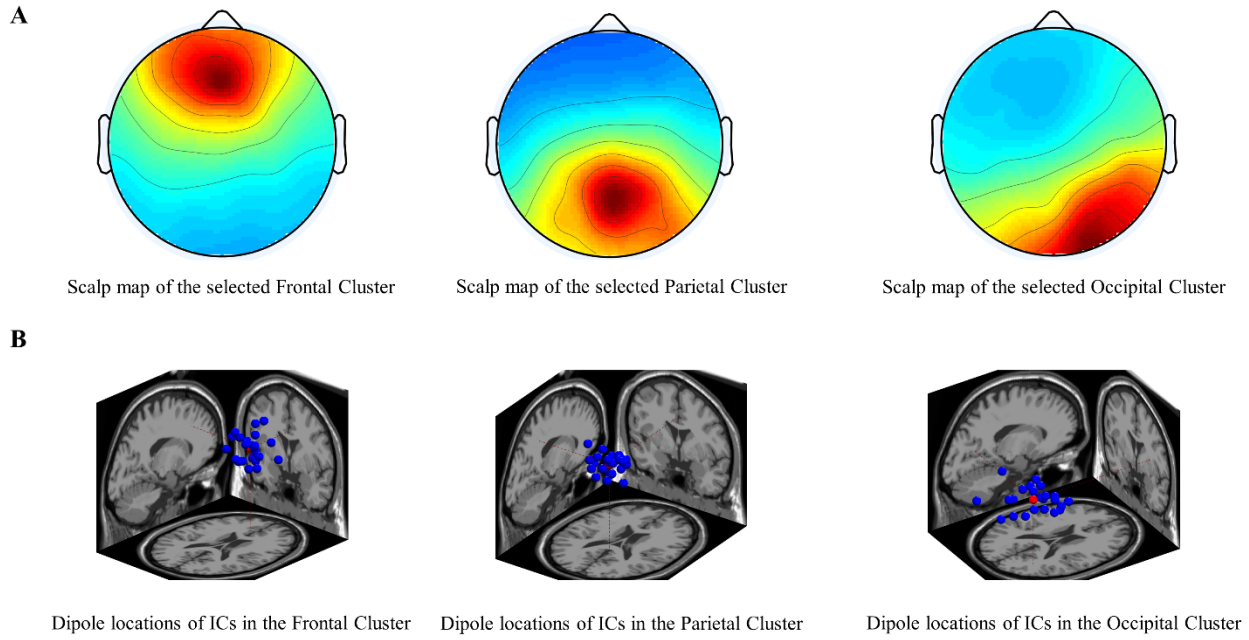


Figure 3. Frontal [Talairach coordinate: (-1, 41, 27)], Parietal [Talairach coordinate: (4, -51, 39)] and Occipital [Talairach coordinate: (30, -70, 15)] clusters selected in the tracking task (A) spatial scalp maps; (B) dipole source locations.

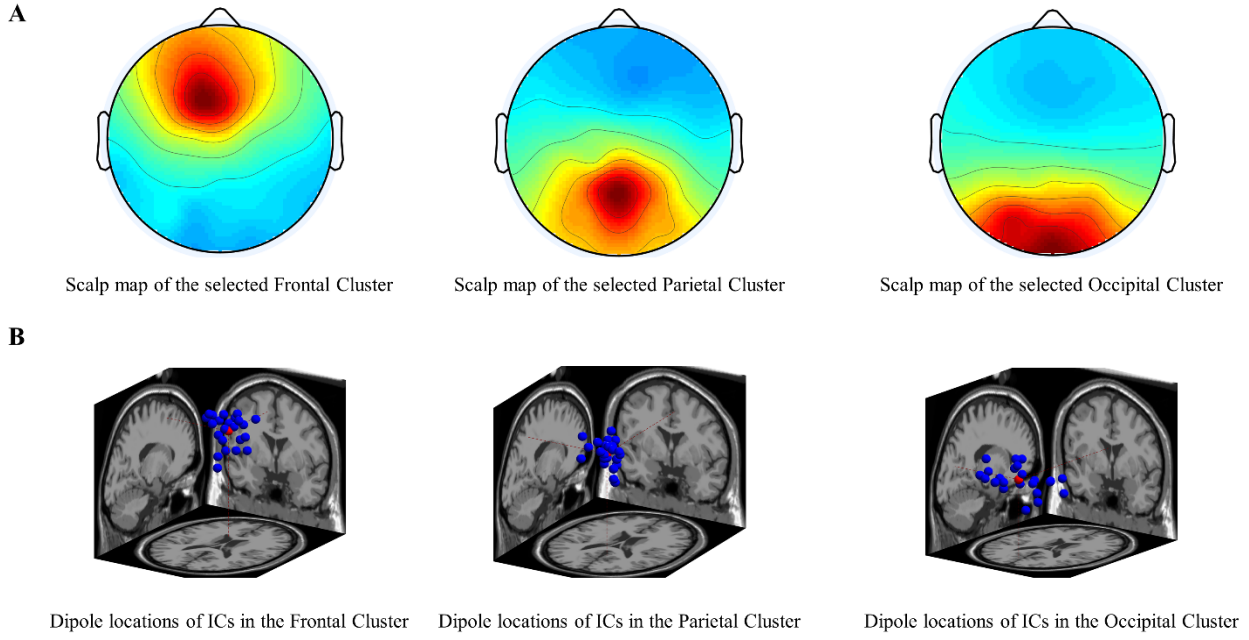
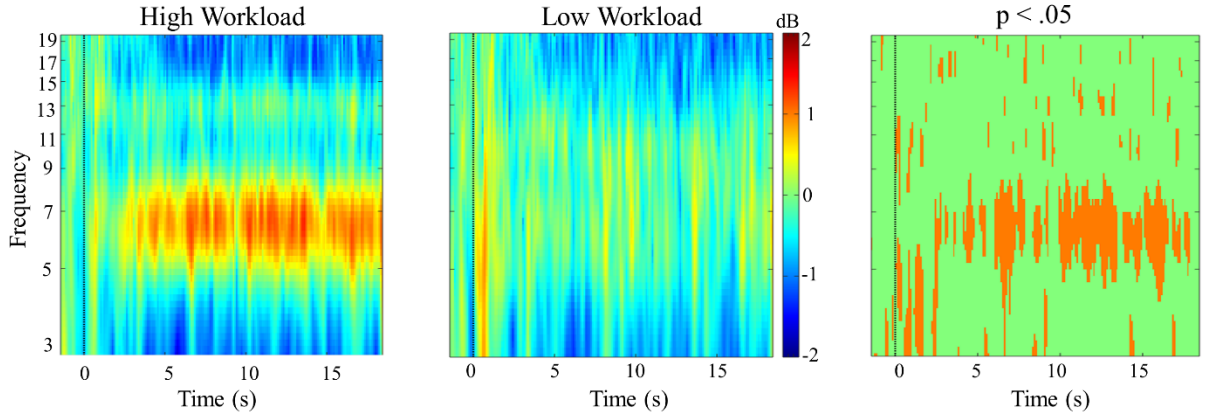


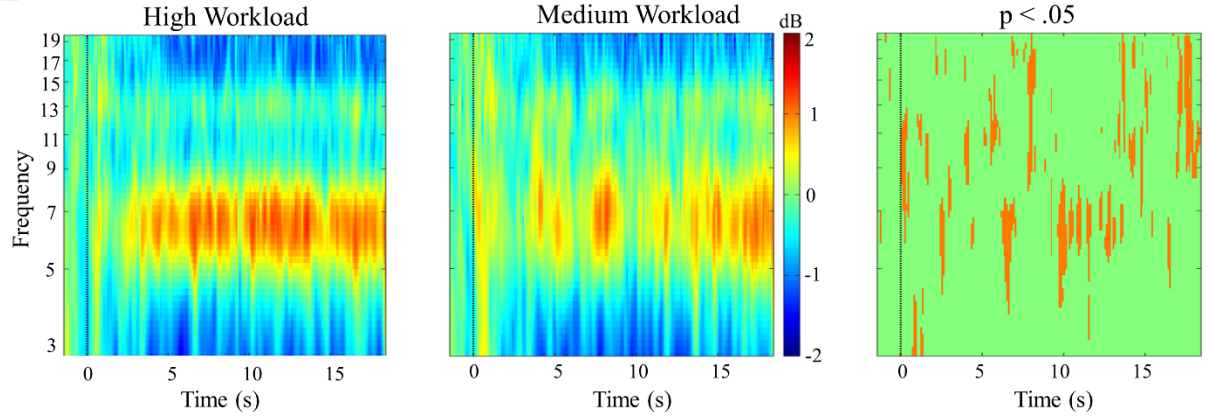
Figure 4. Frontal [Talairach coordinate: (-10, 17, 46)], Parietal [Talairach coordinate: (5, -47, 47)] and Occipital [Talairach Coordinate: (-3, -69, 20)] clusters selected in the collision prediction task (A) spatial scalp maps; (B) dipole source locations.

A

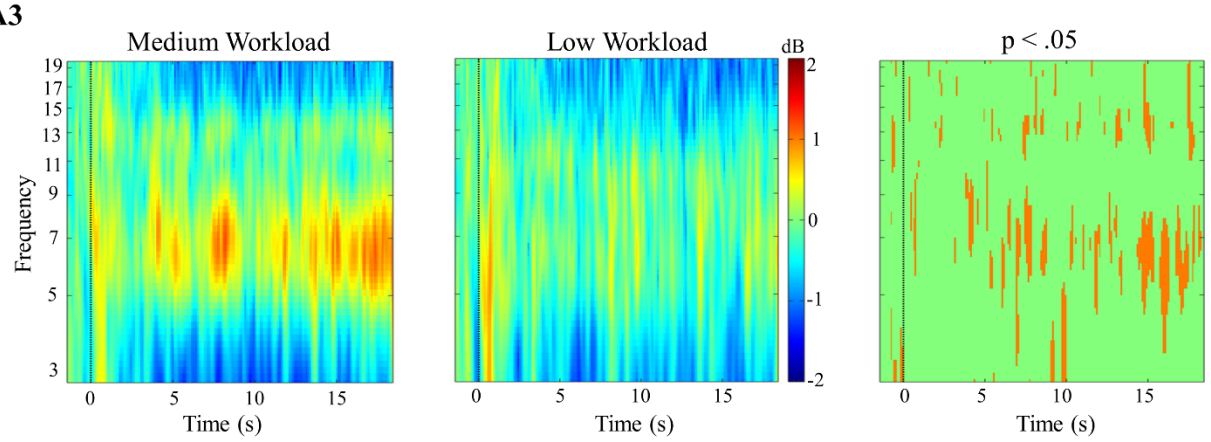
A1


High Workload
Low Workload
dB
p < .05
Frequency
Time (s)

A2


High Workload
Medium Workload
dB
p < .05
Frequency
Time (s)

A3


Medium Workload
Low Workload
dB
p < .05
Frequency
Time (s)

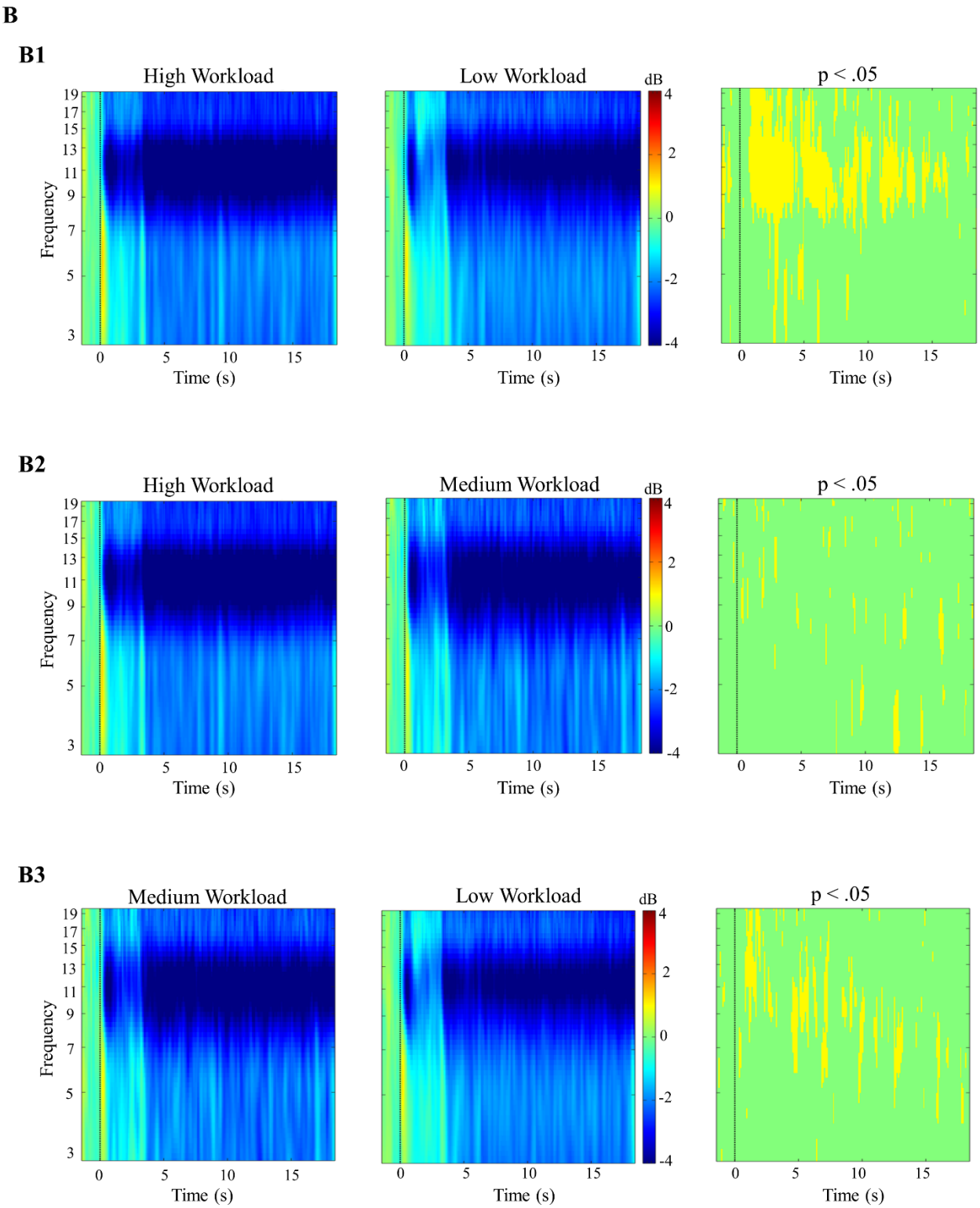
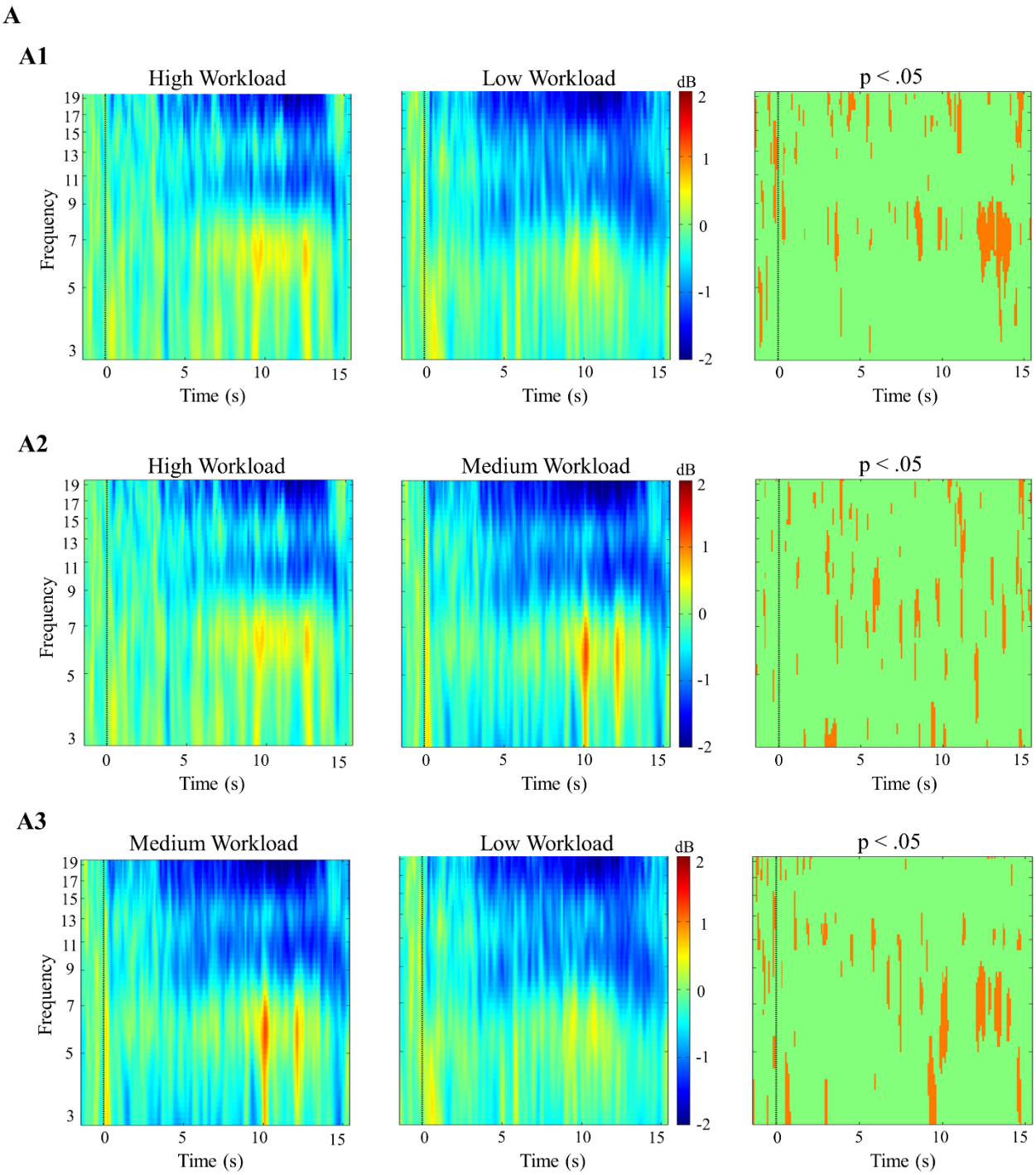


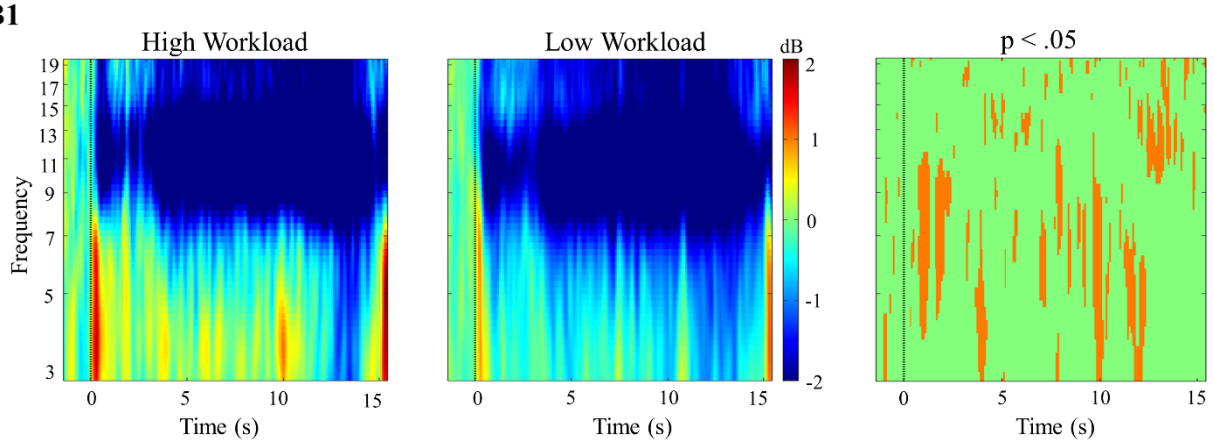
Figure 5: ERSP changes during the tracking task at the (A) Frontal and (B) Occipital Cluster. (A1) shows the ERSP changes at the frontal cluster during high (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant difference between high and low workload conditions ($p < .05$). (A2) shows the ERSP changes at the frontal cluster during high (first panel) and medium (second panel) workload conditions and the third panel shows the statistically significant difference between high and medium workload conditions ($p <$

.05). (A3) shows the ERSP changes at the frontal cluster during medium (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant difference between medium and low workload conditions ($p < .05$). (B1) shows the ERSP changes at the occipital cluster during high (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant difference between high and low workload conditions ($p < .05$). (B2) shows the ERSP changes at the occipital cluster during high (first panel) and medium (second panel) workload conditions and the third panel shows the statistically significant difference between high and medium workload conditions ($p < .05$). (B3) shows the ERSP changes at the occipital cluster during medium (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant difference between medium and low workload conditions ($p < .05$).

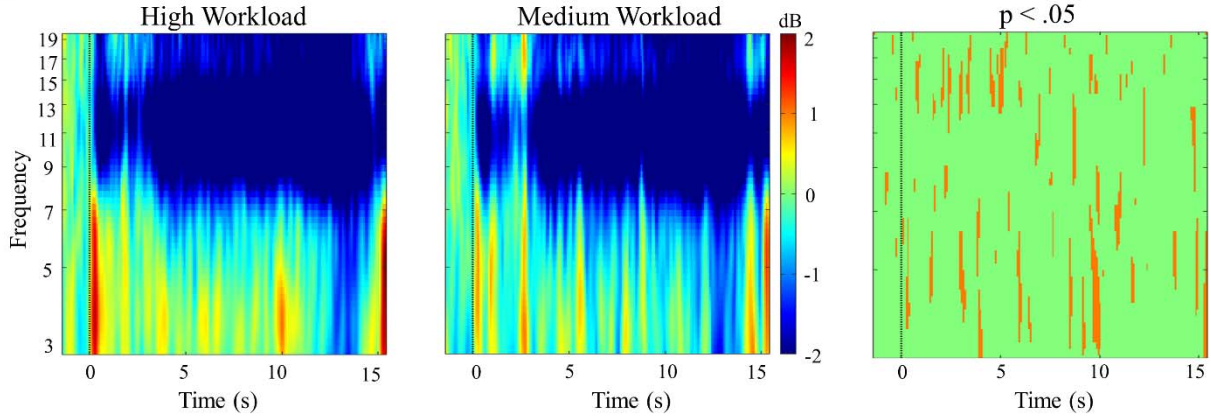


B

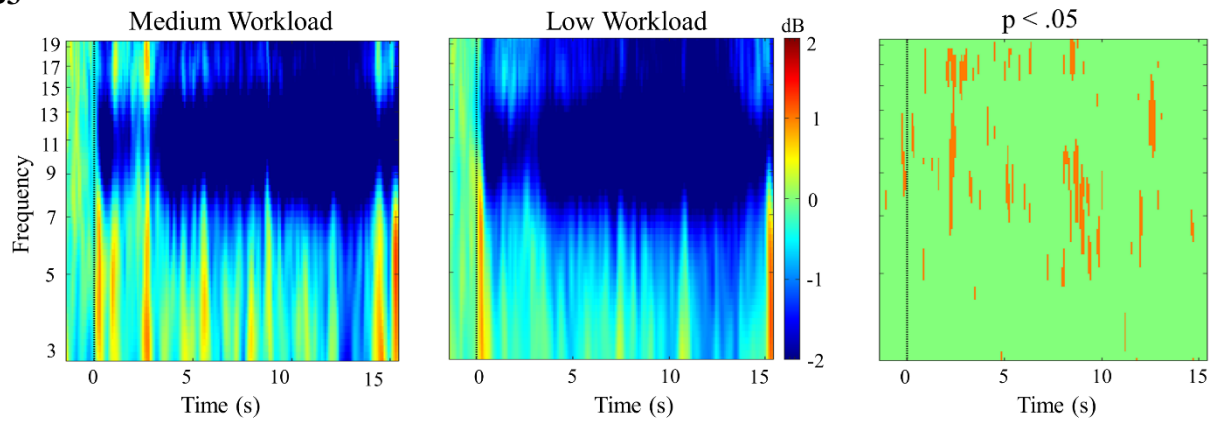
B1


High Workload
Low Workload
dB
p < .05
Frequency
Time (s)

B2


High Workload
Medium Workload
dB
p < .05
Frequency
Time (s)

B3


Medium Workload
Low Workload
dB
p < .05
Frequency
Time (s)

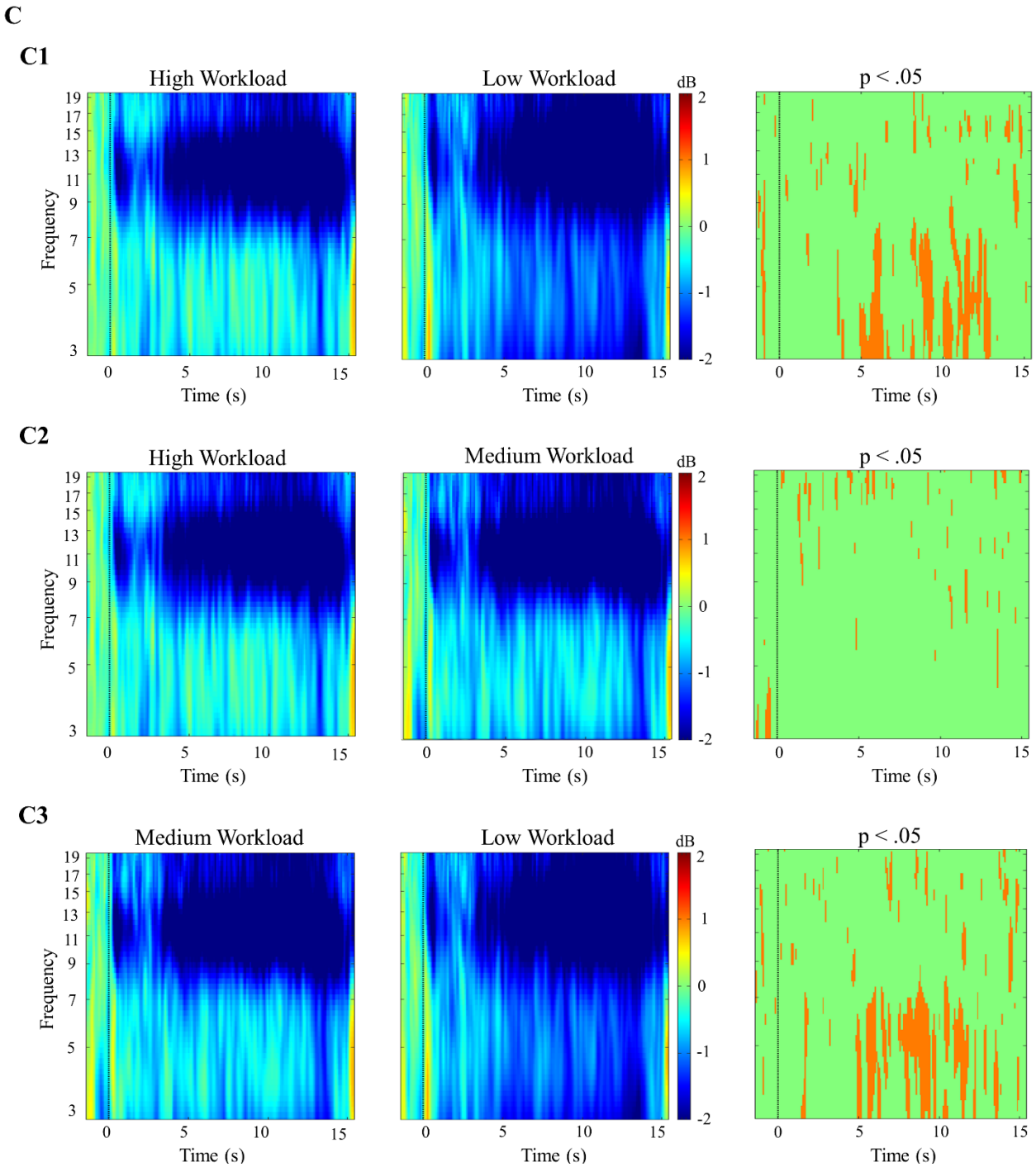


Figure 6: ERSP changes during the collision prediction task at the (A) Frontal, (B) Parietal, (C) Occipital Cluster. (A1) shows the ERSP changes at the frontal cluster during high (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant difference between high and low workload conditions ($p < .05$). (A2) shows the ERSP changes at the frontal cluster during high (first panel) and medium (second panel) workload conditions and the third panel shows the statistically significant difference between high and medium workload conditions ($p < .05$). (A3) shows the ERSP changes at the frontal cluster during medium (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant difference between medium and low workload conditions ($p < .05$). (B1) shows the ERSP changes at the parietal cluster during high (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant

difference between high and low workload conditions ($p < .05$). (B2) shows the ERSP changes at the parietal cluster during high (first panel) and medium (second panel) workload conditions and the third panel shows the statistically significant difference between high and medium workload conditions ($p < .05$). (B3) shows the ERSP changes at the parietal cluster during medium (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant difference between medium and low workload conditions ($p < .05$). (C1) shows the ERSP changes at the occipital cluster during high (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant difference between high and low workload conditions ($p < .05$). (C2) shows the ERSP changes at the occipital cluster during high (first panel) and medium (second panel) workload conditions and the third panel shows the statistically significant difference between high and medium workload conditions ($p < .05$). (C3) shows the ERSP changes at the occipital cluster during medium (first panel) and low (second panel) workload conditions and the third panel shows the statistically significant difference between medium and low workload conditions ($p < .05$).

Figure 6 illustrates the frontal, parietal and occipital clusters' ERSP changes for three workload conditions in the collision prediction task. Statistical analysis on ERSP changes of the frontal cluster showed a significant increase in theta power during high workload condition as compared to low workload condition (Figure 6(A1)). The frontal power during high workload condition was also significantly greater than that of medium workload, as shown in Figure 6(A2). Further, Figure 6(A3) shows that there was a significant increase in the frontal theta power during medium workload as compared to low workload condition in the collision prediction task. The ERSP changes at the parietal cluster (Figure 6(B1)) revealed a significant increase in the theta power in high workload as compared to low workload condition ($p < .05$) and a significant decrease in the alpha power ($p < .05$). Figure 6(B2) shows that there was a significant increase in the theta power and significant decrease in the alpha power at the parietal cluster during the high workload condition as compared to the medium workload condition. In the medium workload condition, the parietal theta power was significantly higher than the low workload condition while the parietal alpha power was significantly lower than that in the low workload condition (Figure 6(B3)). The ERSP changes at the occipital cluster (Figure 6(C1)) revealed a significant increase in the delta and theta power in the high workload as compared to the low workload condition ($p < .05$). Figure 6(C2) shows that there was a significant increase in the delta and theta power at the occipital cluster during the high workload condition as compared to the medium workload condition. In the medium workload condition, the occipital delta and theta power was significantly higher than the low workload condition (Figure 6(C3)).

Power Spectral Density Changes with Mental Workload

Figure 7(A1) illustrates the frontal cluster's ICs' spectral power for low, medium and high workload conditions of the tracking task. The results of the one-way repeated-measures ANOVA showed that frontal theta power varied significantly across workload conditions [$F(2, 46) = 50.931$, $p < .001$, $\eta_p^2 = .822$]. Frontal theta PSD was higher during the high workload as compared to the low workload ($p < .001$) and medium workload condition ($p < .001$). The medium workload condition has significantly higher frontal theta PSD as compared to the low workload condition ($p = .006$). As shown in Figure 7(A2), the results of one-way repeated-measures ANOVA showed that occipital alpha power varied significantly across workload conditions [$F(2, 46) = 24.780$, $p < .001$, $\eta_p^2 = .693$]. Occipital alpha PSD was lower during high workload as compared to low workload ($p < .001$) and medium workload condition ($p = .005$). The medium workload condition had significantly lower occipital alpha PSD as compared to low workload condition ($p = .018$).

For the collision prediction task, the frontal cluster's ICs showed significant variation in theta power across different workload conditions according to the one-way repeated-measures ANOVA [$F(2, 46) = 8.570$, $p = .001$, $\eta_p^2 = .271$]. As shown in Figure 7(B1), frontal theta PSD was higher during high workload as compared to low workload ($p = .002$); however, it was not significantly greater than that of the medium workload condition ($p = .051$). The increase in frontal theta power during the medium workload condition was also not statistically significant as compared to the low workload condition ($p = .336$). However, the parietal cluster's IC's spectral power for low, medium and high workload conditions showed a significant increase in the theta frequency bands and a significant decrease in the alpha band, as shown in Figure 7(B2) and Figure 7(B3).

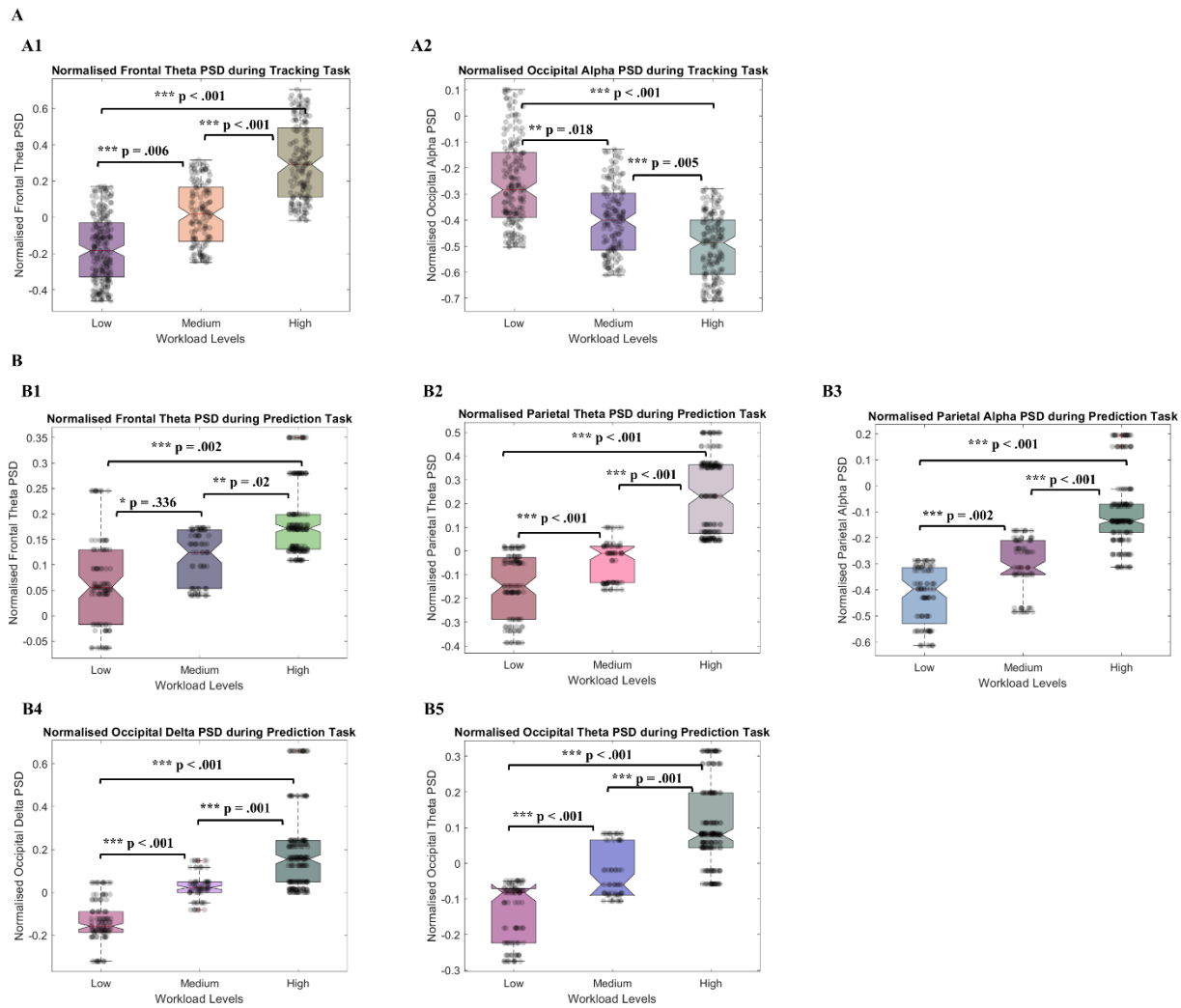


Figure 7: (A) Normalized Power Spectral Density at the Frontal and Occipital ICs selected in the Frontal and Occipital clusters for the tracking task. (A1) shows the normalised frontal theta PSD in the low, medium, and high workload conditions. (A2) shows the normalised occipital alpha PSD for low, medium, and high workload condition for the tracking task. (B) shows the normalized Power Spectral Density at the Frontal, Parietal and Occipital ICs selected in the Frontal, Parietal and Occipital cluster for the collision prediction task. (B1) shows the mean frontal theta PSD in the low, medium, and high workload conditions. (B2) shows the mean parietal theta PSD for the three levels of workload. (B3) shows the mean parietal alpha power for different workload conditions and (B4) shows the

mean occipital delta PSD in the low, medium, and high workload conditions. (B5) shows the mean occipital theta PSD for the three levels of workload condition in collision prediction task.

One-way repeated-measures ANOVA results showed that parietal theta [$F(2, 46) = 47.764$, $p < .001$, $\eta_p^2 = .675$] and alpha [$F(2, 46) = 38.639$, $p < .001$, $\eta_p^2 = .627$] power varied significantly across workload conditions. Parietal theta PSD during low workload was significantly lower than medium ($p < .001$) and high workload conditions ($p < .001$) and the medium workload had lower parietal theta PSD as compared to high workload condition ($p < .001$). There was also a significant decrease in the parietal alpha PSD during the medium ($p = .002$) and the high workload conditions ($p < .001$) as compared to the low workload condition. The parietal alpha PSD was significantly less in the high workload condition as compared to the medium workload ($p < .001$). One-way repeated-measures ANOVA results showed that occipital delta [$F(1.563, 35.951) = 35.321$, $p < .001$, $\eta_p^2 = .606$] and theta [$F(2, 46) = 39.101$, $p < .001$, $\eta_p^2 = .630$] power varied significantly across workload conditions. The occipital delta PSD was significantly higher during the medium ($p < .001$) and the high workload ($p < .001$) as compared to the low workload. Further, occipital delta PSD during the high workload condition was significantly higher than the medium workload condition ($p = .001$). Occipital theta PSD during low workload was significantly lower than the medium ($p < .001$) and high workload conditions ($p < .001$) and the medium workload had lower occipital theta PSD as compared to the high workload condition ($p = .001$).

Eye activity changes with mental workload

As shown in Figure 8(A), pupil size increased with the increasing workload for both tracking and collision prediction tasks. For the tracking task, there was a significant change in the pupil size for different workload conditions as shown by one-way repeated-measures ANOVA [$F(2, 38) = 13.205$, $p < .001$,$\eta_p^2 = .410$]. There was a significant increase in the pupil size for the medium ($p = .0001$) and the high workload condition ($p = .001$) as compared to the low workload condition. However, the increase was not statistically significant for the high workload as compared to the medium workload condition ($p = .313$) in the tracking task.

The results of one-way repeated measures ANOVA shows that in the collision prediction task, there was a significant change in the pupil size for different workload conditions [$F(2, 46) = 9.276$, $p < .001$, $\eta_p^2 = .287$]. There was a significant increase in the pupil size for the medium ($p = .011$) and the high workload condition ($p < .001$) as compared to the low workload condition and no significant increase in pupil size for high workload as compared to the medium workload condition ($p = .180$).

The number of blinks during tracking and collision prediction tasks decreased with the increasing workload, as shown in Figure 8(B). One-way repeated-measure ANOVA was conducted to study the effect of workload variations on the number of blinks, which revealed significant variations in the number of blinks during the tracking task for different workload conditions [$F(2, 46) = 3.624$, $p = .035$, $\eta_p^2 = .136$]. The number of blinks in the low workload condition of the tracking task was significantly greater than that of high workload condition ($p = .015$) but not significantly greater than that of medium workload ($p = .328$). There was no significant decrease in the number of blinks in the high workload condition as compared to medium workload ($p = .106$).

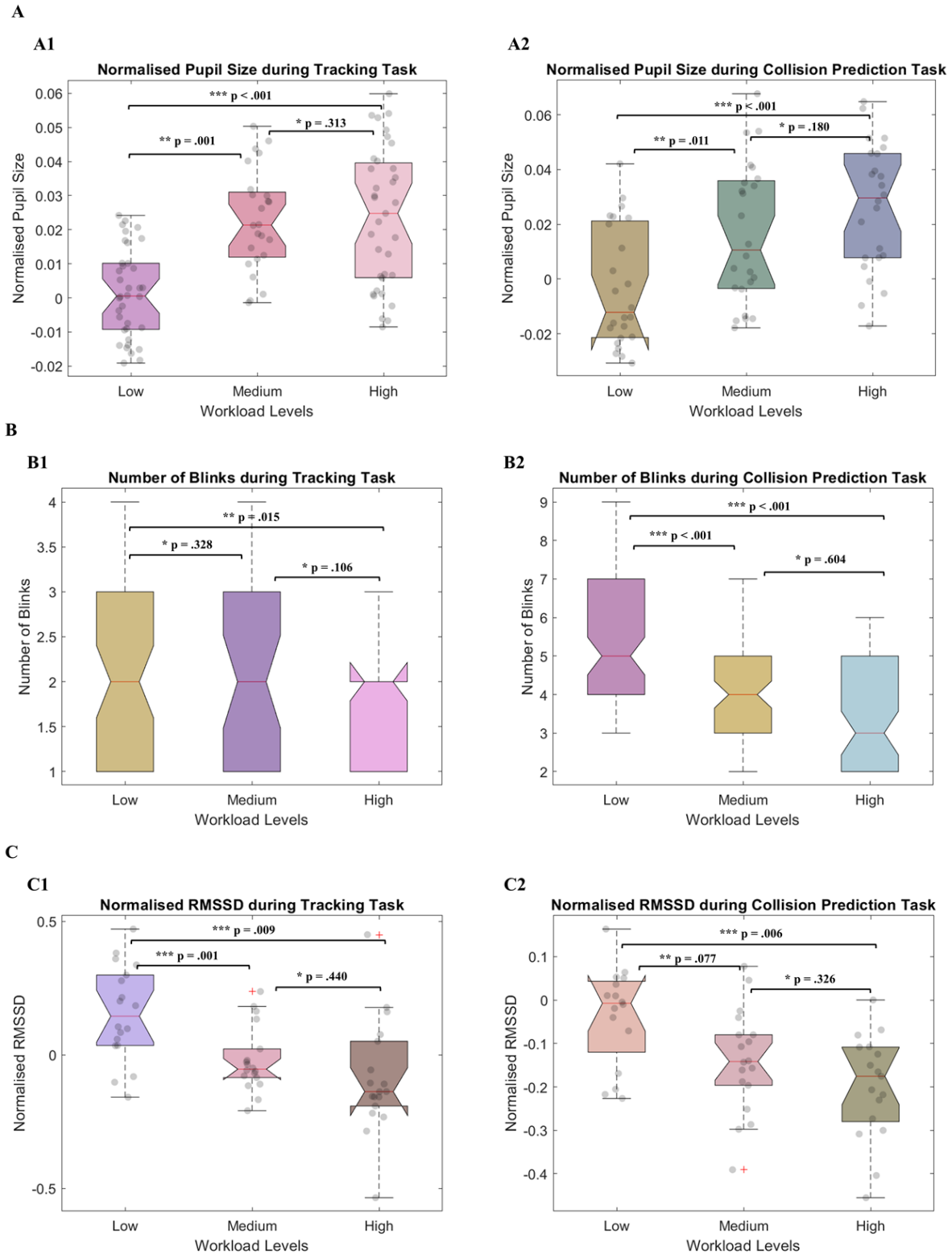


Figure 8: (A) shows the normalized pupil size of all the participants shows a positive trend with the increasing workload. (A1) Normalised pupil size in the three workload conditions of the tracking task. (A2) Normalised pupil size during low, medium, and high workload conditions for the collision prediction task. (B) shows the negative trend in the number of blinks with the increasing workload. (B1) Number of blinks during different workload

conditions of the tracking task. (B2) Number of blinks during the collision prediction task decreases with increasing level of workload. (C) shows the declining trend in the normalized RMSSD of all the participants with the increasing workload. (C1) Normalised RMSSD all the participants in the low, medium, and high workload conditions of the tracking task. (C2) Normalised RMSSD during collision prediction task for the three levels of workload.

The effect of workload on the number of blinks in the collision prediction task was analysed using one-way repeated-measure ANOVA. It showed a significant variation in the number of blinks [$F(2, 46) = 18.586$, $p < .001$, $\eta_p^2 = .447$]. In the low workload condition, the number of blinks was significantly greater in the medium ($p < .001$) and the high workload conditions ($p < .001$). However, the number of blinks in the medium workload condition was not significantly higher when compared to the high workload condition ($p = .604$).

Heart Rate Variability (RMSSD) changes with Mental Workload

Figure 8(C) shows the RMSSD variation for different workload conditions in the tracking and collision prediction task. For the tracking task, there was a significant change in the RMSSD for the different workload conditions, as shown by the one-way repeated-measures ANOVA [$F(2, 34) = 10.171$, $p < .001$, $\eta_p^2 = .374$]. There was a significant decrease in the RMSSD for the medium ($p = .001$) and the high workload condition ($p = .009$) as compared to the low workload condition. The RMSSD during medium and high workload of the tracking task was not significantly different ($p = .440$). Results from one-way repeated-measures ANOVA shows that in the collision prediction task, there was a significant change in the RMSSD for different workload conditions [$F(2, 44) = 4.279$, $p = .022$, $\eta_p^2 = .201$]. RMSSD during the low workload condition was significantly greater than the medium ($p = .077$) and high workload conditions ($p = .006$) of the collision prediction task. There was no significant variation in RMSSD for the medium and high workload conditions ($p = .326$) of the collision prediction task.

Multiple Regression Results

Multiple regression was carried out to investigate whether EEG, eye activity and HRV metrics of workload could significantly predict the performance in the tracking task. The results of the regression indicated that the model explained 54.3% of the variance and that the model was a significant predictor of the tracking performance, $F(3, 67) = 26.543$, $p < .001$. While EEG metrics ($B = .067$, $p = .001$) and eye activity ($B = -.089$, $p < .001$) contributed significantly to the model, HRV metrics did not ($B = -.152$, $p = .125$). The final predictive model was:

$$Performance\ in\ tracking\ task = 0.725 - 0.067 * EEG\ metrics - 0.089 * Eye\ related\ metrics - 0.152 * HRV\ metrics \quad (4)$$

In order to determine whether EEG, eye activity and HRV metrics could significantly predict the performance in collision prediction task, we conducted multiple regression analysis. The results of the regression indicated that the model explained 61.7% of the variance and that the model was a significant predictor of the performance in the collision prediction task, $F(3, 68) = 24.324$, $p < .001$. While eye activity ($B = -.276$, $p = .02$) and EEG metrics ($B = -.532$, $p < .001$) contributed significantly to the model, HRV metrics did not ($B = .444$, $p = .443$). The final predictive model was:

$$Performance\ in\ Collision\ Prediction\ task = 0.055 - 0.532 * EEG\ metrics - 0.276 * Eye\ related\ metrics + 0.444 * HRV\ metrics \quad (5)$$

Discussion

In this study, we designed two simplified tasks based on ATC: tracking and collision prediction tasks. Although both these tasks represent the basic tasks that ATC operators routinely perform, we considered them separately to untangle the differences in the physiological response to workload variations in these tasks.

In order to study workload effects of increasing air traffic, the mental workload in both these tasks was manipulated by varying the number of dots. It was observed that the performance in the tracking task, assessed by the tracking accuracy, degrades significantly with the increasing workload. Similarly, for the collision detection task, the time before collision decreased significantly and collision prediction miss proportion also significantly increased with increasing levels of workload. Together, this indicated an overall decrease in performance with the increasing levels of workload for the collision detection task as participants took longer to identify collisions and were less accurate in identifying collisions when workload increased. Hence, we can confirm that the workload manipulation (by varying the number of dots) in both tracking and collision prediction tasks was successful in eliciting significant performance variations (H1).

In order to assess the mental workload, EEG, eye activity and BVP data were recorded while the participants performed the tasks. For all participants, the component data was disentangled from the scalp EEG signal using independent component analysis. Significant correlations between mental workload and the spectral powers of frontal, parietal and occipital clusters were successfully elucidated.

The tracking task demands allocation of attentional resources to keep track of one, three or five tracking dots moving randomly among distractor dots. Working memory load is sensitive to increased allocation of attentional resources and is reflected by increases in frontal theta power (Klimesch et al., 1998; Klimesch, 1999; Gevins and Smith, 2000). In the tracking task, we observed an increase in the frontal theta power, which confirms that increased working memory load was experienced with increasing workload levels. Tracking dots moving among distractor dots also entails working memory mechanisms related to relevant item maintenance and increases in the memory load. This working memory mechanism was reflected by a decrease in the alpha power (Gevins et al., 1997; Wilson, 2002; Capilla et al., 2014 and Puma et al., 2018). The alpha power is also known to decrease with increased memory load (Fournier et al., 1999; Smith et al., 2001; Fairclough et al., 2005; Ryu and Myung, 2005; Fairclough and Venables, 2006) and task difficulty (Sterman and Mann, 1995; Ota et al., 1996). Our findings also substantiate this working memory mechanism as the occipital alpha power decreases with increasing workload levels in the tracking task.

In the collision prediction task, anticipating the trajectory of the dots and predicting whether the dots would collide requires attention and internal concentration. Delta power is an indicator of attention or internal concentration in mental tasks, and it has been reported to increase with the increase in workload (Sterman and Mann, 1995; Harmony et al., 1996; Wilson, 2002). Our

results demonstrate an increase in the delta power at the occipital sites, which validates that there is an increased allocation of attentional resources with increasing levels of workload in the collision prediction task. Additionally, keeping a tab on the trajectory of six, 12 or 18 eight dots adds to the memory load in the participants. Several studies have shown that theta power is correlated with memory load (Jensen and Tesche, 2002; Jacobs et al., 2006) and working memory capacity (Klimesch, 1996; Klimesch, 1999; Sauseng et al., 2010). In collision prediction task, our results reveal a significant increase in the theta power at the frontal, parietal and occipital clusters, confirming an increase in memory load with increasing levels of workload. Furthermore, our results indicate that with increasing levels of workload, there is a crease in parietal alpha power. This observed alpha band desynchronisation with the increasing workload is related to relevant item maintenance in the working memory (Sterman and Mann, 1995; Gevins et al., 1997; Wilson, 2002; Capilla et al., 2014; Puma et al., 2018) and is known to decrease with increased memory load (Fournier et al., 1999; Smith et al., 2001; Fairclough et al., 2005; Ryu and Myung, 2005; Fairclough and Venables, 2006) and task difficulty (Sterman and Mann, 1995; Ota et al., 1996). However, in the collision prediction task, the most significant decrease in the parietal alpha power was observed a few seconds before the collision. It might be related to the increase in the experienced time pressure (Slobounov et al., 2000) as the participants attempt to identify and click on the colliding pair of dots before the collision happens.

We also explored eye-related metrics and HRV metrics during workload variations. Eye activity data was transformed to pupil size and blink rate information. Pupil size increased significantly with the increasing workload in both tracking and collision prediction tasks. The number of blinks also reduced considerably with the increasing workload in both tasks. Pupil size is a reliable measure of workload (Beatty, 1982, Marquart, 2015, Marquart and Winter, 2015, Mandrick et al., 2016) as it dilates with increasing workload (Batmaz and Ozturk, 2008, Kosch et al., 2018, Truschzinski et al., 2018, Bernhardt et al., 2019, Kearney et al., 2019, Marinescu et al., 2018, Wanyan et al., 2014). Recarte et al., 2008 show that blink inhibition happens in higher workload conditions and so, the blink rate is inversely correlated with the attentional levels and workload experienced by the operator (Veltman and Gaillard, 1996, Brookings et al., 1996, Wilson, 2002, Borghini et al., 2014, Widyanti et al., 2017, Wanyan et al., 2018). RMSSD was found to be negatively correlated with the mental workload in both tasks. This decrease in RMSSD with the increasing workload is widely reported in the literature (Mehler et al., 2011, Cinaz et al., 2010, Cinaz et al., 2013, Parsinejad et al., 2014, Fallahi et al., 2016, Heine et al., 2017, Tjolleng et al., 2017).

Our results show that EEG power spectra at the frontal, parietal and occipital areas, eye activity and HRV metrics can reliably and accurately assess the mental workload of the participants in both tasks. Hence, our second hypothesis (H2) is proved to be true for both tracking and collision prediction tasks. Relating to our third hypothesis (H3), the multiple regression results showed that the performance in the tracking and collision prediction tasks could be predicted from the EEG, eye related and HRV metrics.

Even though EEG, eye activity and HRV measures were able to differentiate between low and high levels of workload sensitively, some of these measures were not able to accurately discern the medium workload from low/high workload conditions. There are two possible reasons for this incoherence reported in the literature: experiment design issue (Kramer, 1991) or inter-

individual differences (Beatty, 1977; Beatty and O'Hanlon, 1979 and Valentino et al., 1993). In our experimental design, the medium workload condition might have required nearly comparable cognitive resources and hence, not qualifying for a significant variation from the low/high workload condition. However, our results showed a significant drop in the performance with increasing workload levels in both the tracking and collision prediction tasks.

Therefore, it is more plausible to reason that this incoherence might be due to the influence of inter-individual differences. It is well understood that the relationship between workload and task demand is not straightforward (Athènes et al., 2002; Chatterji and Sridhar, 2001). Sperandio, 1971 claims that the relationship can be better understood by investigating the strategies employed by human operators to manage their cognitive resources and workload and many researchers agree with this view (Cullen, 1999, Athènes et al., 2002, Averty et al., 2002, Hilburn, 2004, Majumdar et al., 2004). Different participants might reflect workload variations differently based on their cognitive resources and the strategies that they employ for performing the tasks.

Our results also indicate that even though eye activity and HRV metrics are sensitive to task load variations, they may not provide any valuable information on the task that causes the variations in workload. However, the EEG measures were found to be not just sensitive to the workload variations but also the task type. The increases in workload in the tracking task was reflected by the increase in frontal theta power and decrease in occipital alpha power. No significant changes were observed in the parietal theta, alpha, occipital delta, or theta power with the increasing workload in the tracking task. In the collision prediction task, the increase in workload was correlated with the increases in frontal theta, parietal theta, occipital delta and theta power and a decrease in parietal alpha power. No significant variation was observed in the occipital alpha power during the collision prediction task. The neurometrics correlated with the variations in the workload of tracking and collision prediction tasks are different, which proves that our fourth hypothesis (H4) is true. Therefore, neurometrics can help identify the task contributing to the increase in workload in complex ATC environments at a time instant and define the strategies that can be used by the workload adaptive system to mitigate this increase. These results provide evidence that the use of EEG measures in a closed-loop adaptive system can not only aid the decision of 'when' but also 'what' form of automation to deploy to mitigate the workload variations in operators. Hence, the results presented here contribute to the development of adaptive strategies essential for the design of intelligent closed-loop mental workload adaptive ATC systems.

While we have systematically studied the effect of workload variations, the main limitation of this study is that different variables were controlled for the purpose of elucidating the impact of workload variations based on the differences in traffic load in the basic ATC tasks. Our experiment scenario was not very realistic as several environmental factors contribute to the workload experienced by ATC even while they perform these basic tasks. Another limitation is that a prior gaming experience can influence the strategies employed by the participants, thereby, significantly affect the experienced workload. We did not explore the inter-individual differences between participants in this study. Finally, this is a work in progress. We are in the process of building an intelligent mental workload adaptive closed-loop system based on the reported preliminary results where the workload mitigation strategy employed at any time instant will be decided based on the real-time neurometrics of the operator.

Conclusion

The performance and efficiency of a system can be improved by maintaining the operator's workload in the optimal range. In order to elucidate the impact of basic task load variations that comprise the load variations in complex ATC tasks, we separately designed two basic ATC tasks: tracking and collision prediction tasks. EEG spectral power, eye and HRV correlates to mental workload variations for tracking and collision prediction tasks of air traffic controllers are successfully unravelled and provide a comprehensive picture of the workload demands in ATC tasks. Our results demonstrate that EEG, eye and HRV metrics can provide a sensitive and reliable measure to predict the mental workload and performance of the operator. The differences in neural response to increased workload in the tracking and collision prediction task indicate that these neural measures are sensitive to variations and type of mental workload and their potential utility in not just deciding 'when' but also 'what' to adapt, aiding the development of intelligent closed-loop mental workload aware systems. This investigation of physiological indices of workload variation in the basic ATC tasks has applicability to the design of future adaptive systems that integrate neurometrics in deciding the form of automation to be used to mitigate the variations in workload in complex ATC systems.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Acknowledgements

This work was supported in part by the Australian Research Council (ARC) under Discovery Grant DP180100670 and Discovery Grant DP180100656; in part by the Australia Defence Innovation Hub under Contract P18-650825; U.S. Office of Naval Research Global through Cooperative Agreement under Grant ONRG-NICOP-N62909-19-1-2058; and in part by the NSW Defence Innovation Network and NSW State Government of Australia under Grant DINPP2019 S1-03/09.

References

Ahlstrom, Ulf. "An eye for the air traffic controller workload." In *Journal if the Transportation Research Forum*, vol. 46, no. 3. 2010.

Aricò, Pietro, et al. "Adaptive automation triggered by EEG-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment." *Frontiers in human neuroscience* 10 (2016): 539.

Aricò, Pietro, et al. "Human factors and neurophysiological metrics in air traffic control: a critical review." *IEEE reviews in biomedical engineering* 10 (2017): 250-263.

Aricò, Pietro, et al. "How neurophysiological measures can be used to enhance the evaluation of remote tower solutions." *Frontiers in Human Neuroscience* 13 (2019).

Athènes, Sylvie, et al. "ATC complexity and controller workload: Trying to bridge the gap." *Proceedings of the International Conference on HCI in Aeronautics*. AAAI Press Cambridge, MA, 2002.
Averty, Philippe, et al. "Evaluating a new index of mental workload in real ATC situation using psychophysiological measures." *Proceedings. The 21st Digital Avionics Systems Conference*. Vol. 2. IEEE, 2002.
Batmaz, Ihsan, and Mustafa Ozturk. "Using pupil diameter changes for measuring mental workload under mental processing." *JApSc* 8, no. 1 (2008): 68-76.
Beatty, Jackson. "Learned regulation of alpha and theta frequency activity in the human electroencephalogram." *Biofeedback: Theory and research* (1977): 251-370.
Beatty, Jackson, and James F. O'Hanlon. "Operant control of posterior theta rhythms and vigilance performance: Repeated treatments and transfer of training." *Progress in operant conditioning of physiological events* (1979): 235-246.
Beatty, Jackson. "Task-evoked pupillary responses, processing load, and the structure of processing resources." *Psychological bulletin* 91.2 (1982): 276.
Bernhardt, Kyle A., et al. "The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment." *Applied ergonomics* 77 (2019): 83-91.
Blankertz, Benjamin, et al. "The Berlin brain-computer interface: progress beyond communication and control." *Frontiers in Neuroscience* 10 (2016): 530.
Boksem, Maarten AS, Theo F. Meijman, and Monicque M. Lorist. "Effects of mental fatigue on attention: an ERP study." *Cognitive brain research* 25, no. 1 (2005): 107-116.
Borghini, Gianluca, et al. "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness." *Neuroscience & Biobehavioral Reviews* 44 (2014): 58-75.
Borragan Pedraz, Guillermo. "Behavioural bases and functional dynamics of cognitive fatigue." (2016).
Brennan, S. D. "An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder." *Portsmouth: DRA Maritime Command and Control Division. DRA Technical Memorandum (CAD5)* 92017 (1992).
Brookings, Jeffrey B., Glenn F. Wilson, and Carolyne R. Swain. "Psychophysiological responses to changes in workload during simulated air traffic control." *Biological psychology* 42.3 (1996): 361-377.
Byrne, Evan A., and Raja Parasuraman. "Psychophysiology and adaptive automation." *Biological psychology* 42.3 (1996): 249-268.
Calabrese, Edward J. "Dose-response features of neuroprotective agents: an integrative summary." *Critical reviews in toxicology* 38.4 (2008): 253-348.
Capilla, Almudena, et al. "Dissociated α-band modulations in the dorsal and ventral visual pathways in visuospatial attention and perception." *Cerebral Cortex* 24.2 (2014): 550-561.
Casali, John G., and Walter W. Wierwille. "On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues." *Ergonomics* 27.10 (1984): 1033-1050.
Charles, Rebecca L., and Jim Nixon. "Measuring mental workload using physiological measures: a systematic review." *Applied ergonomics* 74 (2019): 221-232.

Chatrian, G. E., E. Lettich, and P. L. Nelson. "Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities." *American Journal of EEG technology* 25, no. 2 (1985): 83-92.
Chatterji, Gano, and Banavar Sridhar. "Measures for air traffic controller workload prediction." *1st AIAA, Aircraft, Technology Integration, and Operations Forum*. 2001.
Cinaz, Burcu, Roberto La Marca, Bert Arnrich, and Gerhard Tröster. "Monitoring of mental workload levels." (2010): 189-193.
Cinaz, Burcu, Bert Arnrich, Roberto La Marca, and Gerhard Tröster. "Monitoring of mental workload levels during an everyday life office-work scenario." *Personal and ubiquitous computing* 17, no. 2 (2013): 229-239.
Cullen, L. "Validation of a methodology for predicting performance and workload (European Experimental Centre Note No. 7/99)." *Brussels, Belgium: EUROCONTROL* (1999).
Cummings, Mary L., and Chris Tsonis. "Deconstructing complexity in air traffic control." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 49. No. 1. Sage CA: Los Angeles, CA: Sage Publications, 2005.
Delorme, Arnaud, and Scott Makeig. "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis." *Journal of neuroscience methods* 134, no. 1 (2004): 9-21.
Delorme, Arnaud, Tim Mullen, Christian Kothe, Zeynep Akalin Acar, Nima Bigdely-Shamlo, Andrey Vankov, and Scott Makeig. "EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing." *Computational intelligence and neuroscience* 2011 (2011).
De Waard, Dick, and Rijksuniversiteit te Groningen. *The measurement of drivers' mental workload*. Netherlands: Groningen University, Traffic Research Center, 1996.
de Waard, Dick, and Ben Lewis-Evans. "Self-report scales alone cannot capture mental workload." *Cognition, technology & work* 16.3 (2014): 303-305.
Di Flumeri, Gianluca, Gianluca Borghini, Pietro Aricò, Nicolina Sciaraffa, Paola Lanzi, Simone Pozzi, Valeria Vignali et al. "EEG-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings." *Frontiers in human neuroscience* 12 (2018): 509.
Di Flumeri, Gianluca, et al. "The dry revolution: evaluation of three different EEG dry electrode types in terms of signal spectral features, mental states classification and usability." *Sensors* 19.6 (2019): 1365.
Do, Tien-Thong, Chin-Teng Lin, and Klaus Gramann. "Human retrosplenial theta and alpha modulation in active spatial navigation." *bioRxiv* (2020).
Fritz, Thomas, et al. "Using psycho-physiological measures to assess task difficulty in software development." *Proceedings of the 36th international conference on software engineering*. 2014.
Gale, Anthony, Robin Davies, and Adrian Smallbone. "EEG correlates of signal rate, time in task and individual differences in reaction time during a five-stage sustained attention task." *Ergonomics* 20.4 (1977): 363-376.
Gevins, Alan, et al. "Monitoring working memory load during computer-based tasks with EEG pattern recognition methods." *Human factors* 40.1 (1998): 79-91.
Gevins, Alan, and Michael E. Smith. "Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style." *Cerebral cortex* 10.9 (2000): 829-839.
Gevins, Alan S., et al. "Effects of prolonged mental work on functional brain topography." *Electroencephalography and Clinical Neurophysiology* 76.4 (1990): 339-350.

Gevins, Alan, Michael E. Smith, Linda McEvoy, and Daphne Yu. "High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice." *Cerebral cortex (New York, NY: 1991)* 7, no. 4 (1997): 374-385.
Gevins, Alan, and Michael E. Smith. "Neurophysiological measures of cognitive workload during human-computer interaction." *Theoretical Issues in Ergonomics Science* 4.1-2 (2003): 113-131.
Gronlund, Scott D., et al. "Role of memory in air traffic control." *Journal of Experimental Psychology: Applied* 4.3 (1998): 263.
Fairclough, Stephen H., Louise Venables, and Andrew Tattersall. "The influence of task demand and learning on the psychophysiological response." *International Journal of Psychophysiology* 56, no. 2 (2005): 171-184.
Fairclough, Stephen H., and Louise Venables. "Prediction of subjective states from psychophysiology: A multivariate approach." *Biological psychology* 71, no. 1 (2006): 100-110.
Fallahi, Majid, Majid Motamedzade, Rashid Heidarimoghadam, Ali Reza Soltanian, and Shinji Miyake. "Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study." *Applied ergonomics* 52 (2016): 95-103.
Fournier, Lisa R., Glenn F. Wilson, and Carolyne R. Swain. "Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training." *International Journal of Psychophysiology* 31, no. 2 (1999): 129-145.
Freeman, Frederick G., et al. "Evaluation of an adaptive automation system using three EEG indices with a visual tracking task." *Biological psychology* 50.1 (1999): 61-76.
Hancock, Peter A. "A dynamic model of stress and sustained attention." *Human factors* 31.5 (1989): 519-537.
Hancock, Peter A., and Joel S. Warm. "A dynamic model of stress and sustained attention." *Journal of Human Performance in Extreme Environments* 7.1 (2003): 4.
Hancock, Peter A., and Gerald Matthews. "Workload and performance: Associations, insensitivities, and dissociations." *Human factors* 61.3 (2019): 374-392.
Harmony, Thalía, Thalía Fernández, Juan Silva, Jorge Bernal, Lourdes Díaz-Comas, Alfonso Reyes, Erzsébet Marosi, Mario Rodríguez, and Miguel Rodríguez. "EEG delta activity: an indicator of attention to internal processing during performance of mental tasks." *International journal of psychophysiology* 24, no. 1-2 (1996): 161-171.
Hart, S. G., and L. E. Staveland. "Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv. Psychol. 52 (C), 139–183 (1988)."
Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." *Journal of the royal statistical society. series c (applied statistics)* 28, no. 1 (1979): 100-108.
Hilburn, Brian. "Cognitive complexity in air traffic control: A literature review." *EEC note* 4.04 (2004).
Heine, Tobias, Gustavo Lenis, Patrick Reichensperger, Tobias Beran, Olaf Doessel, and Barbara Deml. "Electrocardiographic features for the measurement of drivers' mental workload." *Applied ergonomics* 61 (2017): 31-43.
Hjortskov, Nis, et al. "The effect of mental stress on heart rate variability and blood pressure during computer work." *European journal of applied physiology* 92.1-2 (2004): 84-89.
Holm, A., et al. "Estimating brain load from the EEG. Sci. World J. 9, 639–651." (2009).

Innes, Reilly J., Nathan J. Evans, Zachary L. Howard, Ami Eidels, and Scott D. Brown. "A broader application of the detection response task to cognitive tasks and online environments." *Human Factors* (2019): 0018720820936800.
Jacobs, Joshua, Grace Hwang, Tim Curran, and Michael J. Kahana. "EEG oscillations and recognition memory: theta correlates of memory retrieval and decision making." *Neuroimage* 32, no. 2 (2006): 978-987.
Jordan, C. S. "Experimental study of the effect of an instantaneous self assessment workload recorder on task performance." *Report No. DRA/TM (CAD5)/92011. Farnborough: Defence Evaluation & Research Agency* (1992).
Jensen, Ole, and Claudia D. Tesche. "Frontal theta activity in humans increases with memory load in a working memory task." *European journal of Neuroscience* 15, no. 8 (2002): 1395-1399.
Kaber, David B., and Mica R. Endsley. "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task." *Theoretical Issues in Ergonomics Science* 5, no. 2 (2004): 113-153.
Kahneman, Daniel. *Attention and effort*. Vol. 1063. Englewood Cliffs, NJ: Prentice-Hall, 1973.
Kamath, Markad V., and Ernest L. Fallen. "Power spectral analysis of heart rate variability: a noninvasive signature of cardiac autonomic function." *Critical reviews in biomedical engineering* 21.3 (1993): 245-311.
Kassner, Moritz, William Patera, and Andreas Bulling. "Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction." In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*, pp. 1151-1160. 2014.
Ke, Yufeng, et al. "An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task." *Frontiers in human neuroscience* 8 (2014): 703.
Kearney, Peter, et al. "The impact of alerting designs on air traffic controller's eye movement patterns and situation awareness." *Ergonomics* 62.2 (2019): 305-318.
Keil, Andreas, Matthias M. Müller, Thomas Gruber, Christian Wienbruch, Margarita Stolarova, and Thomas Elbert. "Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials." *Clinical neurophysiology* 112, no. 11 (2001): 2057-2068.
Kirwan, Barry, et al. "Human factors in the ATM system design life cycle." *FAA/Eurocontrol ATM R&D Seminar, Paris, France*. 1997.
Klimesch, Wolfgang. "Memory processes described as brain oscillations in the EEG-alpha and theta bands." *Psycoloquy* (1995).
Klimesch, Wolfgang. "Memory processes, brain oscillations and EEG synchronization." *International journal of psychophysiology* 24, no. 1-2 (1996): 61-100.
Klimesch, Wolfgang, et al. "Induced alpha band power changes in the human EEG and attention." *Neuroscience letters* 244.2 (1998): 73-76.
Klimesch, Wolfgang. "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis." *Brain research reviews* 29, no. 2-3 (1999): 169-195.
Kompier, Michiel AJ, and Tage S. Kristensen. "Organizational work stress interventions in a theoretical, methodological and practical context." (2001).

Kosch, Thomas, Mariam Hassib, Daniel Buschek, and Albrecht Schmidt. "Look into my eyes: using pupil dilation to estimate mental workload for task complexity adaptation." In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1-6. 2018.
Kothe, Christian. "Lab streaming layer (LSL)." *https://github. com/sccn/labstreaminglayer. Accessed on October* 26 (2014): 2015.
Kramer, Arthur F. "Physiological metrics of mental workload: A review of recent progress." *Multiple-task performance* (1991): 279-328.
Kramer, Arthur, and John Spinks. "Capacity views of human information processing." (1991).
Lancaster, Jack L., Marty G. Woldorff, Lawrence M. Parsons, Mario Liotti, Catarina S. Freitas, Lacy Rainey, Peter V. Kochunov, Dan Nickerson, Shawn A. Mikiten, and Peter T. Fox. "Automated Talairach atlas labels for functional brain mapping." *Human brain mapping* 10, no. 3 (2000): 120-131.
Landsbergis, Paul A. "The changing organization of work and the safety and health of working people: a commentary." *Journal of occupational and environmental medicine* 45.1 (2003): 61-72.
Lin, Chin-Teng, and Tien-Thong Nguyen Do. "Direct-Sense Brain-Computer Interfaces and Wearable Computers." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2020).
Lopez-Gordo, Miguel Angel, Daniel Sanchez-Morillo, and F. Pelayo Valle. "Dry EEG electrodes." *Sensors* 14.7 (2014): 12847-12870.
Lin, Chin-Teng, Kuan-Chih Huang, Chih-Feng Chao, Jian-Ann Chen, Tzai-Wen Chiu, Li-Wei Ko, and Tzyy-Ping Jung. "Tonic and phasic EEG and behavioral changes induced by arousing feedback." *NeuroImage* 52, no. 2 (2010): 633-642.
Majumdar, Arnab, et al. "The factors affecting airspace capacity in Europe: A cross-sectional time-series analysis using simulated controller workload data." *The Journal of Navigation* 57.3 (2004): 385.
Makeig, Scott. *Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones*. No. NHRC-92-31. NAVAL HEALTH RESEARCH CENTER SAN DIEGO CA, 1993.
Makeig, Scott, Marissa Westerfield, T-P. Jung, Sonia Enghoff, Jeanne Townsend, Eric Courchesne, and Terrence J. Sejnowski. "Dynamic brain sources of visual evoked responses." *Science* 295, no. 5555 (2002): 690-694.
Mandrick, Kevin, et al. "Neural and psychophysiological correlates of human performance under stress and high mental workload." *Biological psychology* 121 (2016): 62-73.
Marinescu, Adrian Cornelius, Sarah Sharples, Alastair Campbell Ritchie, Tomas Sanchez Lopez, Michael McDowell, and Hervé P. Morvan. "Physiological parameter response to variation of mental workload." *Human factors* 60, no. 1 (2018): 31-56.
Marquart, Gerhard, Christopher Cabrall, and Joost de Winter. "Review of eye-related measures of drivers' mental workload." *Procedia Manufacturing* 3 (2015): 2854-2861.
Marquart, Gerhard, and Joost De Winter. "Workload assessment for mental arithmetic tasks using the task-evoked pupillary response." *PeerJ Computer Science* 1 (2015): e16.
Matousek, Milos, and Ingemar Petersén. "A method for assessing alertness fluctuations from EEG spectra." *Electroencephalography and clinical neurophysiology* 55.1 (1983): 108-113.
Matthews, Gerald, et al. "The psychometrics of mental workload: Multiple measures are sensitive but divergent." *Human Factors* 57.1 (2015): 125-143.
Mehler, Bruce, Bryan Reimer, and Ying Wang. "A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload." (2011).

Mogford, Richard H., et al. *The Complexity Construct in Air Traffic Control: A Review and Synthesis of the Literature*. CTA INC MCKEE CITY NJ, 1995.
Montgomery, Doil D., et al. "Single channel QEEG amplitudes in a bright, normal young adult sample." *Journal of Neurotherapy* 2.4 (1998): 1-7.
Murai, Koji, Yuji Hayashi, and Seiji Inokuchi. "A basic study on teammates' mental workload among ship's bridge team." *IEICE TRANSACTIONS on Information and Systems* 87.6 (2004): 1477-1483.
Murphy, A. A., et al. "245 VALIDATION OF SIMULATION-BASED TRAINING IN NEONATAL RESUSCITATION: USE OF HEART RATE VARIABILITY AS A MARKER FOR MENTAL WORKLOAD." (2004): S122-S122.
Negri, Lucas Hermann. "PeakUtils." *URL: https://pypi. python. org/pypi/PeakUtils (visited on 07/20/2016)* (2018).
Nickel, Peter, and Friedhelm Nachreiner. "Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload." *Human factors* 45.4 (2003): 575-590.
Niosh, N. *The changing organization of work and the safety and health of working people*. Technical Report 2002-116, National Institute for Occupational Safety and Health (NIOSH), 2002.
O'DONNELL, Robert D. "Workload assessment methodology." *Cognitive processes and performance* (1986).
Onton, Julie, Arnaud Delorme, and Scott Makeig. "Frontal midline EEG dynamics during working memory." *Neuroimage* 27, no. 2 (2005): 341-356.
Oostenveld, Robert, and Thom F. Oostendorp. "Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull." *Human brain mapping* 17, no. 3 (2002): 179-192.
Oostenveld, Robert, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. "FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data." Computational intelligence and neuroscience 2011 (2011).
Ota, Toshio, Ryoichi Toyoshima, and Toshio Yamauchi. "Measurements by biphasic changes of the alpha band amplitude as indicators of arousal level." *International journal of psychophysiology* 24, no. 1-2 (1996): 25-37.
PA Van Dongen, Hans, Gregory Belenky, and James M Krueger. "A local, bottom-up perspective on sleep deprivation and neurobehavioral performance." *Current topics in medicinal chemistry* 11.19 (2011): 2414-2422.
Parasuraman, Raja. "Adaptive automation: from theory to practice." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 45. No. 4. Sage CA: Los Angeles, CA: SAGE Publications, 2001.
Parasuraman, Raja, and Matthew Rizzo, eds. *Neuroergonomics: The brain at work*. Oxford University Press, 2008.
Parasuraman, Raja. "Neuroergonomic perspectives on human systems integration: Mental workload, vigilance, adaptive automation, and training." (2015).
Parsinejad, Payam, and Rifat Sipahi. "A touchscreen game to induce mental workload on human subjects." In *2014 40th Annual Northeast Bioengineering Conference (NEBEC)*, pp. 1-2. IEEE, 2014.
Pashler, Harold. "Dual-task interference in simple tasks: data and theory." *Psychological bulletin* 116.2 (1994): 220.

Pfurtscheller, Gert. "EEG event-related desynchronization (ERD) and synchronization (ERS)." *Electroencephalography and Clinical Neurophysiology* 1.103 (1997): 26.
Pion-Tonachini, Luca, Ken Kreutz-Delgado, and Scott Makeig. "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website." *NeuroImage* 198 (2019): 181-197.
Popovic, Djordje, et al. "Sensitive, diagnostic and multifaceted mental workload classifier (PHYSIOPRINT)." *International Conference on Augmented Cognition*. Springer, Cham, 2015.
Prinzel, Lawrence J., Frederick G. Freeman, Mark W. Scerbo, Peter J. Mikulka, and Alan T. Pope. "A closed-loop system for examining psychophysiological measures for adaptive task allocation." *The International journal of aviation psychology* 10, no. 4 (2000): 393-410.
Privitera, Claudio M., Laura W. Renninger, Thom Carney, Stanley Klein, and Mario Aguilar. "Pupil dilation during visual target detection." *Journal of Vision* 10, no. 10 (2010): 3-3.
Puma, Sébastien, et al. "Using theta and alpha band power to assess cognitive workload in multitasking environments." *International Journal of Psychophysiology* 123 (2018): 111-120.
Radüntz, Thea. "Signal quality evaluation of emerging EEG devices." *Frontiers in physiology* 9 (2018): 98.
Radüntz, Thea. "Dual frequency head maps: A new method for indexing mental workload continuously during execution of cognitive tasks." *Frontiers in physiology* 8 (2017): 1019.
Radüntz, Thea, and Beate Meffert. "User experience of 7 mobile electroencephalography devices: comparative study." *JMIR mHealth and uHealth* 7.9 (2019): e14474.
Reason, James. "Human error: models and management." *Bmj* 320.7237 (2000): 768-770.
Recarte, Miguel Ángel, et al. "Mental workload and visual impairment: Differences between pupil, blink, and subjective rating." *The Spanish journal of psychology* 11.2 (2008): 374.
Reid, G. B. N., E. Thomas, and Hancock PAEM. "Najmedin (Ed), The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. Human mental workload.. Human mental workload." *Advances in psychology* 52 (1988): 185-218.
Rodgers, Mark D., and Gena K. Drechsler. *Conversion of the CTA, Inc., en route operations concepts database into a formal sentence outline job task taxonomy*. No. DOT/FAA/AM-93/1. Civil Aerospace Medical Institute, 1993.
Rouse, William B. "Adaptive aiding for human/computer control." *Human factors* 30, no. 4 (1988): 431-443.
Ryu, Kilseop, and Rohae Myung. "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic." *International Journal of Industrial Ergonomics* 35, no. 11 (2005): 991-1009.
Sauseng, Paul, Birgit Griesmayr, Roman Freunberger, and Wolfgang Klimesch. "Control mechanisms in working memory: a possible function of EEG theta oscillations." *Neuroscience & Biobehavioral Reviews* 34, no. 7 (2010): 1015-1022.
Scerbo, Mark W. "Theoretical perspectives on adaptive automation." (1996): 37-63.
Scerbo, Mark W. *The efficacy of psychophysiological measures for implementing adaptive technology*. NASA Langley Research Center, 2001.
Schmorrow, Dylan, Kay M. Stanney, Glenn Wilson, and Peter Young. "Augmented cognition in human–system interaction." *Handbook of human factors and ergonomics* (2006): 1364-1383.
Shou, Guofa, Lei Ding, and Deepika Dasari. "Probing neural activations from continuous EEG in a real-world task: time-frequency independent component analysis." *Journal of neuroscience methods* 209.1 (2012): 22-34.

Slobounov, S. M., K. Fukada, R. Simon, M. Rearick, and W. Ray. "Neurophysiological and behavioral indices of time pressure effects on visuomotor task performance." *Cognitive Brain Research* 9, no. 3 (2000): 287-298.
Smith, Michael E., Alan Gevins, Halle Brown, Arati Karnik, and Robert Du. "Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction." *Human Factors* 43, no. 3 (2001): 366-380.
Sperandio, J. C. "Variation of operator's strategies and regulating effects on workload." *Ergonomics* 14.5 (1971): 571-577.
Sterman, M. B., and C. A. Mann. "Concepts and applications of EEG analysis in aviation performance evaluation." *Biological psychology* 40.1-2 (1995): 115-130.
Tjolleng, Amir, Kihyo Jung, Wongi Hong, Wonsup Lee, Baekhee Lee, Heecheon You, Joonwoo Son, and Seikwon Park. "Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals." *Applied ergonomics* 59 (2017): 326-332.
Truschzinski, Martina, et al. "Emotional and cognitive influences in air traffic controller tasks: An investigation using a virtual environment?." *Applied ergonomics* 69 (2018): 1-9.
Ullsperger, P., A-M. Metz, and H-G. Gille. "The P300 component of the event-related brain potential and mental effort." *Ergonomics* 31, no. 8 (1988): 1127-1137.
Valentino, Dominic A., J. E. Arruda, and S. M. Gold. "Comparison of QEEG and response accuracy in good vs poorer performers during a vigilance task." *International Journal of Psychophysiology* 15, no. 2 (1993): 123-133.
Van Acker, Bram B., et al. "Understanding mental workload: from a clarifying concept analysis toward an implementable framework." *Cognition, Technology & Work* 20.3 (2018): 351-365.
Veltman, J. A., and A. W. K. Gaillard. "Physiological indices of workload in a simulated flight task." *Biological psychology* 42, no. 3 (1996): 323-342.
Wanyan, Xiaoru, Damin Zhuang, and Huan Zhang. "Improving pilot mental workload evaluation with combined measures." *Bio-medical materials and engineering* 24, no. 6 (2014): 2283-2290.
Wanyan, Xiaoru, Damin Zhuang, Yingzi Lin, Xu Xiao, and Jin-Woo Song. "Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation." *International Journal of Industrial Ergonomics* 64 (2018): 1-7.
Welch, Peter. "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms." IEEE Transactions on audio and electroacoustics 15, no. 2 (1967): 70-73.
Wickens, Christopher D., et al. *Engineering psychology and human performance*. Psychology Press, 2015.
Wickens, Chris, and Pamela S. Tsang. "Workload." (2015).
Widyanti, Ari, Nisha Faradila Sofiani, Herman Rahadian Soetisna, and Khoirul Muslim. "Eye blink rate as a measure of mental workload in a driving task: convergent or divergent with other measures?." *Eye* 8, no. 2 (2017).
Wilson, Glenn F. "An analysis of mental workload in pilots during flight using multiple psychophysiological measures." *The International Journal of Aviation Psychology* 12, no. 1 (2002): 3-18.
Xie, Bin, and Gavriel Salvendy. "Prediction of mental workload in single and multiple tasks environments." *International journal of cognitive ergonomics* 4.3 (2000): 213-242.
Yeh, Yei-Yu, and Christopher D. Wickens. "Dissociation of performance and subjective measures of workload." *Human Factors* 30.1 (1988): 111-120.

Yerkes, R. "M.. & Dodson, JD (1908)." *The relation of strength of stimulus to rapidity of habit forma* (1908).
Zhao, Guozhen, Yong-Jin Liu, and Yuanchun Shi. "Real-time assessment of the cross-task mental workload using physiological measures during anomaly detection." *IEEE Transactions on Human-Machine Systems* 48.2 (2018): 149-160.

Supplementary Material

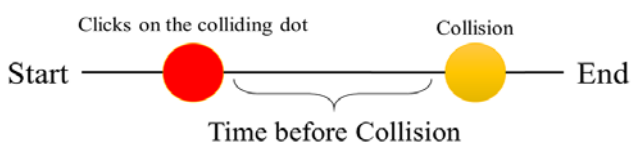


Figure 1. A schematic diagram describing how time before collision was calculated in the collision prediction task

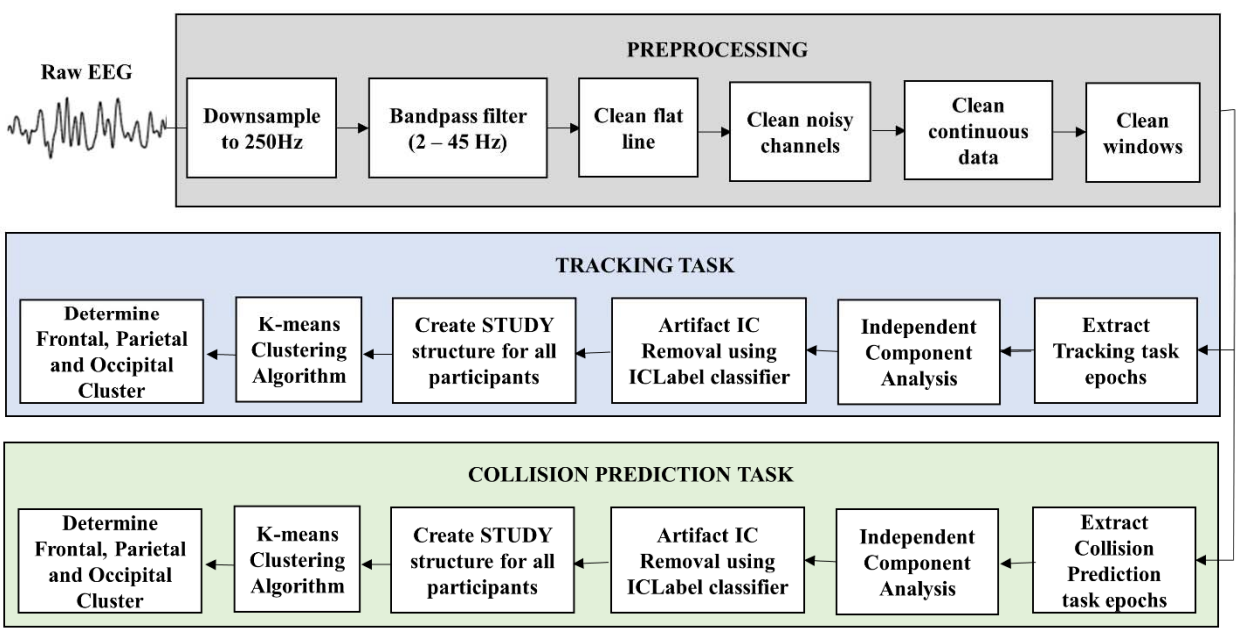


Figure 2. The EEG preprocessing and processing pipeline used for tracking and collision prediction tasks.