7-23-2013

A Quantitative analysis of the mental workload demands of MRAP vehicle drivers using physiological, subjective, and performance assessments

Elizabeth Khol

R·I·T

A Quantitative Analysis of the Mental Workload Demands of MRAP Vehicle Drivers using Physiological, Subjective, and Performance Assessments

by

Elizabeth A. Khol

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Industrial Engineering

Department of Industrial & Systems Engineering
Kate Gleason College of Engineering

Rochester Institute of Technology
Rochester, NY

July 23, 2013

DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING

KATE GLEASON COLLEGE OF ENGINEERING

ROCHESTER INSTITUTE OF TECHNOLOGY

ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

M.S. DEGREE THESIS

The M.S. Degree Thesis of Elizabeth A. Khol has

been examined and approved by the thesis committee as

satisfactory for the thesis requirement for the

Master of Science Degree.

Approved by:

____

Dr. Matthew Marshall, Thesis Advisor
Associate Professor, Industrial & Systems Engineering

____

Dr. Jacqueline Mozrall, Thesis Committee Member
Professor and Associate Dean, College of Engineering

____

Dr. Shannon Foley, Thesis Committee Member
Naval Surface Warfare Center, Dahlgren Division

Abstract

United States Special Operations Command (USSOCOM) Operators and vehicle Commanders are specially trained United States military Warfighters that have the demanding task of operating or working onboard Mine Resistant Ambush Protected (MRAP) All Terrain Vehicles (M-ATVs). Their missions encounter significant mental demands resulting from fatigue, highly stressful situations, and interactions with Government Furnished Equipment (GFE). Excessive mental demands can be the primary factor leading to compromised vehicle communication, missed improvised explosive device (IED) detection, and increased incidents of vehicle roll-over. Research has demonstrated the consequences of mental overloading including increased errors, performance decrements, distraction, cognitive tunneling and inadequate time to appropriately process information. The objectives of this thesis were to evaluate the extent to which task-related factors impact the mental workload of Warfighters and to evaluate the consistency among the three categories of mental workload metrics.

The 14 participants studied in this research were Marine Corps personnel who had heavy vehicle driving experience. Physiological, subjective and performance measures were collected during a four-segment course that progressed in difficulty and analyzed across all participants to assess changes in mental workload. It was found that task-related factors impacted the mental workload of Warfighters. The subjective metric was able to capture changes in workload more accurately than biosignals. Due to technical problems with the biosignal data, comparison of consistency across metrics was inconclusive. The subjective workload ratings were significantly different between course segments and experience levels. The experiment resulted in workload ratings that increased by as much as 94% between segments and were 18% higher among novice drivers. This study showed that mental workload fluctuates while driving in a stressful situation, despite training and experience, and consequently, detection performance will be impacted which could have very adverse consequences. There is the need for additional research to have a better understanding of the true impact of mental workload on MRAP vehicle drivers, especially in an operational environment.

Table of Contents

List of Abbreviations

AFRL	Air Force Research Lab
ANOVA	Analysis of Variance
ARL HRED	Army Research Labs Human Research and Engineering Directorate
ECG	Electrocardiogram
EEG	Electroencephalography
ERP	Event Related Potential
ERSP	Event-Related Spectral Perturbation
FBCB2	Force XXI Battle Command Brigade and Below
Fro-theta	Frontal theta power/activity
GFE	Government Furnished Equipment
GSR	Galvanic Skin Response
HEOG	Horizontal Electrooculogram
HET	Heavy Equipment Transporter
HRV	Heart Rate Variability
HSI	Human Systems Integration
HUMAN	Human Universal Measurement and Assessment Network
ICA	Independent Component Analysis
IED	Improvised Explosive Device
IgA	Immunoglobulin A
LCS	Littoral Combat Ship
LRA	Linear Regression Analysis
MCH	Modified Cooper-Harper scale
MRAP	Mine Resistant Ambush Protected
NASA	National Aeronautics and Space Administration

NSWCDD	Naval Surface Warfare Center Dahlgren Division
OEF	Operation Enduring Freedom
OIF	Operation Iraqi Freedom
OPD	Ohmic perturbation duration
Par-alpha	Parietal alpha power/activity
RMS	Ride Motion Simulator
SWAT	Subjective Workload Assessment Technique
TARDEC	Tank and Automotive Research, Development and Engineering Center
TLX	Task Load Index
UAV	Unmanned Aerial Vehicle
USSOCOM	United States Special Operations Command
VAS-F	Visual Analog Scale for Fatigue
VDT	Visual Display Terminal
VEOG	Vertical Electrooculogram
VSW	Visual-analog Subjective Workload
WP	Workload Profile

List of Tables

List of Equations

List of Figures

Introduction

Warfighters face many difficulties while operating or working onboard Mine Resistant Ambush Protected (MRAP) vehicles. United States Special Operations Command (USSOCOM) Operators and vehicle Commanders are some of the most taxed and most stressed Warfighters conducting missions for the United States military. Their deployments are often long and deep into enemy territory. Their missions have significant mental demands resulting from fatigue and highly stressful situations. Excessive mental demands can be the primary factor leading to compromised vehicle communication, missed improvised explosive device (IED) detection, and increased incidents of vehicle roll-over.

One potential major source of mental workload on MRAP drivers is the Government Furnished Equipment (GFE) subsystems. A few examples of these subsystems include the remote weapons turret, the crosshairs system, and the Force XXI Battle Command Brigade-and-Below (FBCB2). Multiple GFE subsystems are installed inside MRAP vehicles in an attempt to improve situational awareness. It has been demonstrated by many researchers that excessive presentation of information can lead to increased mental workload, and ultimately overload. However, little is known about how much workload from the GFE, or other sources, will compromise the driver's communications and proper IED detection.

The main focus of this study was to evaluate mental workload experienced by MRAP drivers during computer-based, high-fidelity simulated missions. Few studies have been able to collect either operational or simulator data specifically related to an MRAP vehicle. This thesis used data previously obtained from a study that subjected Warfighters to combinations of high and low levels of communications and IED detection while driving through a simulated course. Physiological, subjective and performance measures of mental workload were collected from each Warfighter in order to assess the Warfighter's ability to process information and interpret their surroundings while driving. The focus of this thesis was on mental workload since driving efforts are primarily cognitive and information-processing intensive.

The primary objective of this thesis was to evaluate the extent to which task-related factors impacted the mental workload of Warfighters. A secondary objective was to evaluate the extent to which

the three categories of mental workload metrics demonstrated similar sensitivity levels to the task conditions. Due to the many settings under which workload needs to be measured, it is unlikely that one ideal form of measurement will ever be found for all applications. However, it may be possible to determine the most appropriate measures for a particular situation (Miller, 2001), which this thesis sought to do for MRAP drivers using a controlled, laboratory environment in order to evaluate their feasibility for implementation in an operational environment. The overarching goal was to identify metrics that are reliable, sensitive, and easy to use, and determine whether a simple approach can be used or if an instrument-based approach is required for the desired sensitivity.

The results of this research were intended to provide the data to support or refute the theory that MRAP vehicle drivers experience varying levels of mental workload that can negatively impact IED detection performance. If such a finding is reached, this study also sought to identify which specific mental workload assessment tool(s) may be the most accurate, conclusive, and easy to use outside of a laboratory environment.

Literature Review

Mental Workload

Mental workload is a complex topic with no widely accepted definition or way to measure. Hancock and Meshkati (1988) defined mental workload as “the operator’s evaluation of the attentional load margin (between their motivated capacity and the current task demands) while achieving adequate task performance in a mission-relevant context.” Other researchers defined mental workload as the “portion of operator information processing capacity or resources that is actually required to meet system demands” (Eggemeier, Wilson, Kramer, and Damos, 1991). Both definitions acknowledge the consumption of limited cognitive resources required by a system. Individuals have finite resources and it is critical to understand and assess how task demands consume these limited resources (Wickens and Hollands, 2000).

As technology continues to advance, there is the desire to provide these advancements to military Warfighters in order to enhance their preparedness for combat. It is important that these technologies are

assessed by human systems integration (HSI) experts in order to understand the impact these technologies have on the Warfighter, both good and bad. Many current HSI considerations focus heavily on physical human requirements and impacts. There is also a critical need to assess the mental demands that technological advancements place on the Warfighter. These technological advancements generally come in the form of displays that provide enhanced information about the environment, the location of fellow troops, and known or suspected locations of enemy forces. The large number of complex displays can increase the level of information processing required and elevate the risk of overloading. These displays along with the demands of other tasks, such as driving or communication, all compete for the Warfighter's attentional resources, potentially compromising situational awareness and impacting primary task performance (National Research Council, 1997). One tool, the Land Warrior System, intended to help Warfighters by having them wear a head-mounted eye piece that can provide maps and allow them to see around walls, while remaining in safety, through the use of a camera and video feed. During certain missions, soldiers did not use the map on the eye piece screen to identify the location of fellow soldiers. Instead, they handled the situation that was right in front of them, despite the presence of a map. These advancements may have intended to help, but presented visual noise that interfered with the way the Warfighter performed his duties (Shachtman, 2009).

Factors that Impact Mental Workload

Researchers have many reasons to study mental workload. It is a complex topic with no clear solution on exactly what causes an individual to be underloaded or overloaded. Research by Huey and Wickens (1993) categorized workload factors into two broad categories—individual or operator-related and activity or task-related. Within each category are numerous factors that can affect not only the workload level itself, but also how it is perceived. Some examples of operator-related factors include the level of performance exhibited, the level of effort exerted to complete a task, the amount of cognitive processing required, one's perception of the task, the ease or difficulty of the interaction between the operator and the task, and an operator's capabilities and limitations, such as memory, vision, hearing, and physical strength. There are also several examples of task-related factors including performance criteria,

the nature of the task, the precision required, the quality, presentation, and modality of the task, and the characteristics of input or response devices.

Cain (2007) highlighted additional specific factors that can impact mental workload. He noted that excessive task demands can lead to elevated levels of mental workload which could ultimately compromise performance. Additionally, he noted that these high demands could stress an individual’s limited cognitive resources beyond their capacity. Another factor contributing to mental workload is poor task design or system layout. For example, requiring an individual to search displays in a suboptimal manner will result in increased mental workload. Finally, simultaneous tasks competing for the same mental resources can also lead to overloading.

Consequences of Not Considering Mental Workload

Understanding the impacts of mental workload on tasks can also help researchers to recognize the causes and consequences of insufficient task performance. Researchers are well aware that excessive workload can be detrimental to performance. To help ensure optimal performance and satisfaction, individuals want to minimize their workload, and they will do their best to perform at the level dictated by their tasking. However, it is possible that the tasking can become overwhelming. Any increases in task uncertainty or unpredictability can also lead to higher levels of workload by causing stress from time pressures to meet tighter deadlines or working longer hours from overtime. These considerations are important not only for workload, but also for budget constraints and job satisfaction. Ignoring individual differences in mental workload can also impact task performance. Understanding that novices experience higher baseline workload levels than skilled experts can ensure tasks and duties are assigned appropriately to lead to success. For example, a report by Neill (2011) showed that nurses who experience high workload levels compromise patient care by neglecting non-essential care and procedures. As a result, patients may experience reduced satisfaction and nurses will face higher levels of job dissatisfaction and burnout. Therefore, it is up to researchers to determine the best ways to measure workload in order to determine how much workload is too much for certain individuals and different jobs.

It is also important for military vehicles to undergo an assessment of their impact on mental workload. In order to improve Warfighter safety, it is crucial to thoroughly understand how mental workload impacts MRAP vehicle Operators and Commanders. While MRAP vehicles have been shown to help save lives overseas, the consequences of driver overload can be potentially fatal. Between November 2007 and January 2010, there were more than 230 MRAP roll-overs, many of which were directly attributed to driver error (Sanborn, 2010). Additionally, in 2011 alone, there were more than 16,000 makeshift bomb attacks in Afghanistan (Brook, 2012). Therefore, the vigilance of the driver is crucial to properly identify these threats to avoid an incident. These two examples highlight the importance of ensuring MRAP vehicle driver workload does not adversely impact performance.

Relationship between Workload and Performance

Performance can be understood as successfully completing the task at hand without failure or damage to operator or machine (Lysaght, Hill, Dick, Plamondon, Linton, Wierwille, Zaklad, Bittner, Jr., and Wherry, 1989). The Yerkes-Dodson Law characterizes the relationship between workload and performance as an inverted U-shaped function. Essentially, as arousal, or workload, increases, an individual's performance also increases, but only to a certain point. Once this optimum is reached, performance begins to decline as workload continues to increase. Additionally, more complex tasks will have a lower performance optimum and require a lower workload level to reach that optimum (Yerkes and Dodson, 1908). The Yerkes-Dodson Law can further be explained by demonstrating that individuals experiencing low task stress are not fully engaged in their activities and will not improve performance. Under high levels of sustained stress, individuals must develop ways to cope with that stress, which will also ultimately degrade performance. Therefore, a moderate level of stress is best because individuals will be engaged in their tasks and are able to focus their energy to ultimately improve performance (Jamal, 2007).

A study by Cummings and Nehme (2009) sought to develop predictive models of performance and the impact of workload on unmanned aerial vehicle (UAV) operators. It is understood that mental workload increases with larger amounts of information processing imposed on individuals. This

experiment used operator utilization, or the percent of time busy, as an indicator of workload. The findings confirmed the hypothesis that high levels of workload would leave the operator too busy to process information appropriately. Conversely, under-utilized operators overlooked important information because they became complacent and were not fully engaged in the task at hand. The Yerkes-Dodson Law proved to be appropriate to help predict both human and system performance in this situation.

In order to maintain a certain level of performance, individuals develop strategies to cope with workload. They can reschedule, postpone, or even eliminate tasks in order to achieve the desired level of performance. These coping mechanisms are often used when workload levels are too high, but there is also the risk of having workload levels too low. If workload levels are too high, errors increase because individuals are not able to respond fast enough to task requirements. If workload levels are too low, individuals can experience boredom and a lack of vigilance (Huey and Wickens, 1993).

Workload imbalances can occur under many different scenarios. Warfighters are just one group of individuals subject to under- or overloading. At very low levels of workload, the Warfighter can face boredom, reduced alertness and a lack of attention (Brookhuis and de Waard, 2001). Many tasks performed by Warfighters are vigilance tasks and underloading is a serious problem. Understimulation resulting from monitoring for infrequent events should be avoided because lapses in attention can lead to missed cues. High workload levels lead to performance decrements as well as cause distraction, cognitive tunneling or inadequate time to properly process information (Brookhuis, and de Waard, 2001). The optimum level of workload is desired so that Warfighters are not bored but also not overwhelmed, causing performance to suffer. The pace of the Warfighter's environment requires them to process large amounts of information and make critical decisions in a short amount of time.

By establishing a link between workload levels and human performance, several outcomes can be obtained. Manpower and personnel requirements can be established, helping to ensure the proper number of individuals on a task, as well as the appropriate knowledge, skills and abilities for the task. Additionally, the allocation of workload between Warfighters and automation can be established to help ease the burden and ensure optimum performance.

Relationship between Workload and Stress

Stress is understood as a psychological state of tension resulting from a detected threat (Congleton, Jones, Shiflett, McSweeney, and Huchingson, 1997). Sources of stress can come from time pressure, environmental stressors such as noise, heat, vibration, or poor lighting, and psychological factors including anxiety, fatigue, or danger, to name a few. Along with workload, stress has been found to have an interaction with performance, often affecting it in a negative manner. Techniques such as design changes or training attempt to mitigate the negative impacts of stress (Huey and Wickens, 1993).

Many responsibilities placed upon the Warfighter are stressful and these individuals have many other things on their mind beyond just daily work duties. Stress has also been shown to result in excessive task demands and performance decrements (Embrey, Blackett, Marsden and Peachey, 2006). Warfighters subjected to varying levels of physiological stressors such as stress and fatigue consequently experience an increase in mental workload. As systems and demands on the Warfighter become more complex, there is a desire to decrease their stress and workload so they can continue to operate and perform at a high level. Assuming that workload induces stress, one way to identify levels of stress faced by individuals is to measure workload (Congleton, et.al., 1997).

Biron (2012) reported about military efforts that are currently underway by the Air Force Research Laboratory (AFRL) to study how stress affects performance of unmanned aerial vehicle (UAV) pilots. The Human Universal Measurement and Assessment Network (HUMAN) Laboratory is working on integrating a wide variety of sensors, from electroencephalography (EEG) to eye trackers, galvanic skin response, voice stress analyzers, and brainwave monitors, to try and gather the most complete picture of what happens to these pilots as they experience stress. Researchers note that all pilots will experience overloading but at different levels, depending on their level of expertise. The same is likely true of MRAP vehicle drivers. The research initiative at the HUMAN Lab seeks to identify when operators become stressed to help improve system design and manage workload levels.

Mental Workload Assessment Tools

Mental workload measurement devices can be broadly grouped into three distinct categories: physiological, or objective, measures, subjective measures, and performance measures. Physiological measures are objective in that they attempt to capture data without any personal opinion or emotions, subjective measures are based on individual opinions, and performance measures are surrogate measures that evaluate task performance and can be used to vary workload levels. The study performed to collect data for this thesis used all three measurement tools.

There are several factors to consider when selecting metrics to evaluate workload. These factors serve as criteria to help determine which tasks contribute to mental workload (Miller, 2001). According to Wierwille and Eggemeier (1993), the three most important factors to consider when conducting experiments in test and evaluation environments are sensitivity, intrusion, and diagnosticity. The data gathered for this thesis were collected in a six degree-of-freedom motion-based simulator representative of military ground vehicles. Since a simulator is a test and evaluation environment, these three factors were considered most important. Additionally, metrics that rank highly on these three factors will likely face greater acceptance when expanding testing to operational military environments. The military desires metrics that are simple yet robust, providing practical results with minimal impact to participants.

The first metric, sensitivity, desires to discern differences in the workload levels imposed upon an individual (Wierwille and Eggemeier, 1993). Specifically for MRAP vehicles, it is very important to select metrics that are able to assess individual differences in workload either within or between systems. One specific type of sensitivity—global—is an especially important consideration. Global sensitivity means workload consumes multiple types of resources and in varying amounts. A second important factor is intrusion. In order to be able to assess mental workload, first within a light-skinned combat vehicle simulator and in the future in a variety of environments, it is imperative that the workload metric not interfere with an individual's performance in a way that could compromise the performance of the task. The third factor, diagnosticity, is critical because it allows researchers to identify specific aspects of an individual's task that lead to varying levels of workload or provide assignable causes to workload

(Wierwille and Eggemeier, 1993). Diagnosticity allows researchers to understand specific factors that cause workload or types of resources that are being consumed.

The aforementioned factors are the primary areas of interest for this thesis effort. Other additional factors that are important but not considered in this research include validity, reliability, cost, time required, interval of collection, operator acceptance, and ease of data collection, processing, and analysis (Miller, 2001).

Objective/Physiological Tools

Using physiological tools to measure mental workload has many benefits. The results obtained from these tools can be compared across multiple participants, and with high inter-rater reliability because of the standardization of the tools. The inter-rater reliability also allows results to be generalized across a sampling group, or population, depending on the similarity of testing conditions. These tools lend themselves to statistical analysis and interpretation, aiding in the assessment of participants. The NATO Research and Technology Organisation (2001) recommends the use of objective metrics because individuals cannot voluntarily influence the results. Consequently, these tools provide unbiased measures of workload. Finally, the nature of these tools allows researchers to measure workload continuously over time and not just at one particular instance of an individual's experience.

While there are many benefits to using physiological measures, there are also some drawbacks. Certain tools can require cumbersome equipment, such as electroencephalography (EEG). Extensive equipment requirements can be a major limitation to using these tools in an operational environment. For example, in an operational setting, there may be space constraints, mobility requirements, or tasks that require high precision that cannot have equipment interference. Additionally, expert data interpreters, or extensive learning to achieve expertise, may be required, resulting in additional costs as well as a delay in processing the collected data. A consideration to be mindful of from Lee (2000) is that certain measurement devices can be sensitive to other environmental factors that are unrelated to workload which could undermine validity and reliability of results. For example, pupil dilation can be affected by fluctuating light levels or significant changes in screen contrast, which are unrelated to mental workload.

Many studies have been performed using a wide range of physiological tools to assess mental workload in both laboratory and operational environments. However, it is clear that there is no consensus on which measurement tools are best either overall or in a specific environment. It is often up to the researcher to decide what metric should be used since a standard approach is lacking. The lack of consistency when objectively measuring mental workload often makes comparisons across studies very difficult even if the same metric is used.

Appendix A provides a table outlining the results of several studies that used physiological tools. Across 16 studies, 17 different techniques were used to assess workload. Some studies attempted to compare results across different workload measures, while others simply sought to validate one particular technique. Most of the studies were able to show some change in workload but the inconsistency in techniques used and results obtained make it nearly impossible to determine which situations truly had too much workload to the point requiring system redesign. These inconsistencies make comparisons between studies and generalizability to other situations extremely difficult.

Subjective Tools

Subjective mental workload tools are used in an attempt to capture an individual’s personal interpretation of a task’s demands (Cain, 2007). Subjective tools are often used because they are generally very easy to implement regardless of the testing environment, whether in a lab setting or an operational environment, since no complicated instrumentation is required. Subjective tools generally only require a paper form to be filled out which is generally not intrusive or time consuming. This assumes the task has been completed or the paperwork does not interfere. These tools have high face validity, and are cost effective. The NATO Research and Technology Organisation (2001) recommends using subjective measures because they have often been validated in a variety of environments, making them applicable to many different settings. Subjective tools such as the Modified Cooper-Harper scale (MCH), NASA Task Load Index (NASA-TLX), and the Subjective Workload Assessment Technique (SWAT) have all been validated in test and evaluation environments, flight simulators and laboratory studies (NATO Research

and Technology Organisation, 2001). These tools are beneficial for assessing one specific task, but generally not for combinations of tasks that must be performed at the same time.

There are, however, several drawbacks to be mindful of when considering using subjective tools. Given the large individual variability in results, there is very limited inter-rater reliability and the repeatability of the results is limited. The results are subject to individual biases across participants (NATO Research and Technology Organisation, 2001). Although scales are generally predefined and ratings clearly described, subjects will set their own criteria for ratings (Embrey, et. al., 2006). Since the ratings generally produce ordinal data, the ability to perform certain types of statistical parametric analyses with the data may be limited. These tools suffer from limited precision and experimenters may not be able to determine how workload fluctuates during a particular task. Finally, these tools may be potentially disruptive to the task or activity at hand by distracting an individual from their task in order to complete a subjective assessment. Disrupting a task could negatively affect the results, both of the task and the rating, especially if the interruption occurs at a critical point in an activity. Fortunately, under certain laboratory or simulator conditions, it has been shown that subjective ratings can be administered up to 15 minutes after a task has been performed and ratings will not be impacted (NATO Research and Technology Organisation, 2001).

Appendix B contains a summary of several studies that have used subjective measures and the results obtained. Some of these studies used a single subjective measure in combination with either physiological or performance measures. Other studies used several subjective measures to compare the effectiveness or validate a tool. Consistent with the physiological metrics, the results using subjective measures are somewhat inconclusive. NASA-TLX was the most popular metric for these selected studies; however, it did not outperform the other metrics in all cases. Even though the results are inconsistent, these studies show that subjective metrics are effective in assessing an individual’s workload in a variety of laboratory settings. Obtaining reliable subjective ratings across several laboratory settings indicates that in controlled environments, subjective measures are a robust tool. However, further testing should be done in operational settings to determine if similar results can be obtained. Validating subjective

measures across multiple environments would indicate that these tools offer the best way to get meaningful data at the lowest cost. Providing justification for broad implementation of a relatively easy tool would allow for more comparisons to be made across studies because the tool can be tested under many settings without much effort or cost.

Performance Measures

Performance measures rely on primary and secondary tasks to assess mental workload. An underlying assumption to using performance measures is that the increased processing required for higher levels of workload will degrade performance of the tasks being performed. Additional tasks beyond the primary responsibilities of the participant are added to vary the workload level. Accuracy and efficiency of the primary and secondary tasks are used to assess an individual's workload. The drawbacks to performance measure are that they can lead to underloading, overloading, or confounding effects such as training and experience. Additionally, they are very task-specific and not generalizable between tasks (Gawron, 2000). The study performed to collect data for this thesis introduced secondary tasks in order to manipulate workload levels and assess performance.

Environments Used to Study Mental Workload

Studies can be conducted in laboratory settings, simulators, or operational environments. Regardless of the setting that is used, the desire is for the collection environment to be as realistic as possible so results have practical significance and are generalizable to similar environments.

Operational Environments

Collecting data in operational environments allow experimenters to conduct research in the setting where the results are often employed. Operational environments are often very different from the controlled environments of a laboratory. These environments can suffer from noise and no precise ability to duplicate the exact scenario among multiple participants (Carsten and Brookhuis, 2005).

A couple studies have been performed that measured workload in an operational environment. The NATO Research and Technology Organisation (2001) discussed the advantages of the Subjective Workload Analysis Technique (SWAT), NASA Task Load Index (NASA-TLX), and the Bedford Scale

which is derived from the Cooper Harper Scale. It was determined that all three scales were appropriate metrics for use in operational environments. The Littoral Combat Ship (LCS) Manpower, Workload, and Fatigue study collected workload data underway onboard LCS through time on task measurements to identify specific billets at risk for experiencing high workload levels and negative impacts from fatigue and ship motion. At this time, no findings or conclusions have been reported (Gattie, Mead, Baron, Hamilton, Iden, and Thurber, 2011). Due to the limited amount of operational data collected on mental workload, it is clear that more research needs to be performed. The next step for many studies should be to implement the laboratory set-up into an operational environment to validate findings.

Laboratory Environments

Laboratory environments make use of simulators to collect data in an environment that is modeled after the real world. Technological advancements over time have made enhancements to the fidelity of simulators and their representation of real-world environments. The generalizability of laboratory results depends greatly on the fidelity of simulators. Mock-ups or simulators may not entirely represent true workload demands, but they can provide some preliminary results or findings and help direct future studies. The benefit to using laboratory environments is they provide a safe place to test new set-ups, configurations, or technological advancements. They are also controlled environments which can help reduce confounding variables.

Laboratory studies have been conducted and proved the sensitivity of experiments conducted in a simulated setting. A simulation done for guided missile cruisers included an assessment of workload over time (Santoro, Kieras, and Pharmer, 2004). A seven-point rating scale was used to assess workload; however, no training, definitions or instructions were provided. The findings showed subjective workload increases were significant over time even though the variability between individuals' ratings was not very high. Santoro, et.al. (2004) demonstrated that subjective ratings of workload could be linked to the distribution of task activities over time, but each watchstation—areas onboard a Navy ship where specific tasks are performed—was affected differently. Another simulator for the guided missile cruiser command and control environment (Berka, Levendowski, Ramsey, Davis, Lumicao, Stanney, Reeves, Harkness

Regli, Tremoulet, and Stibler, 2005) used real-time electroencephalogram (EEG) in an attempt to demonstrate the possibility of developing a closed-loop system that could reallocate tasks and streamline an individual's cognitive workload. This low-fidelity simulator consisted of a 17-inch monitor displaying images along with a spoken language interface. Even with this simple simulator, varying levels of mental workload and resulting performance could be detected.

While operational environments are often the ideal location to assess mental workload, there are practical limitations to testing in this type of environment. Laboratory environments have been shown to be effective substitutes for operational environments because they can produce authentic results. An automobile driving study by Reimer and Mehler (2011) showed that once a participant had adapted to an environment, their base level of arousal was higher in the real-world environment than a simulator. This is likely due to the fact that the consequences of mistakes in the real-world environment are much more significant than in a simulator. However, the same study credits a medium fidelity simulator for providing a similar pattern of response to an increase in mental demands as was observed for a field condition.

A European driving study conducted by Engström, Johansson, and Östlud, (2005), compared the results obtained from a simulator and field driving experiment through the use of secondary tasks. The goal was to assess the generalizability of results between environments. The results found consistent trends between both environments. However, the physiological workload was higher in the field environment, indicating an increased effort because of the actual risk associated with real-world driving. It is important to conduct additional research on simulator set-ups to better understand the effectiveness of simulators as compared to operational environments.

Signal Detection Theory

Detection tasks require individuals to continuously monitor for infrequent signal events. Individuals can lose their attentiveness by monitoring an event for a long period of time or from distracters that divert their attention away from the signal. There is an underlying assumption that vigilance can be sustained, even when signals are infrequent. However, if the mental workload becomes too much for an individual, people lose their ability to focus on the task (Craig, 1985).

Signal detection tasks require that an operator be able to detect that a signal is present along with the inevitable noise. Signal detection theory can be used when a subject is presented with a task and their job is to respond either "yes" a signal was present or "no" a signal was not present. From these responses come four possible outcomes based on signal presentation and the subject's response. These are shown below in Table 1.

	State of World		
Operator Behavior		Signal Present (+ Noise)	Signal Absent (Noise only)
	"Yes" (Signal seen)	Hit (H)	False Alarm (FA)
	"No" (No signal perceived)	Miss (M)	Correct Rejection (CR)

Table 1: Outcomes of Signal Detection Theory

Two important parameters that can be evaluated with signal detection theory are sensitivity, also known as d', and response criterion, also known as β. Sensitivity represents the separation of internal noise from a signal and noise; the inherent ability of observer to discriminate signal from noise. Low sensitivity leads to high misses and false alarms since the item of interest is easily confused with noise. High sensitivity leads to very few misses and false alarms because the item of interest is easily distinguished as either a signal or noise.

The response criterion represents the decision strategy of the observer and the likelihood of an individual to indicate if there is a target or a distracter. When the observer has a conservative criterion, classified by β greater than one, they require a large amount of evidence to indicate a signal is present. Therefore, the probability of a correct rejection is higher than the probability of a hit. Risky criterion with β less than one, indicates that less evidence is required for the observer to identify a signal is present. In this case, the probability of a hit is larger than the probability of a correct rejection. Finally, observers can be neutral in their criterion, as shown when the probability of a correct rejection is equal to the probability of a hit (Wickens, Lee, Liu, and Becker, 2004). Rewards or punishments can be assigned to certain outcomes in an attempt to alter participant behavior (Knoblauch and Maloney, 2012).

The costs and benefits associated with each of the four signal detection theory outcomes can affect the response criterion. If the observer cannot afford to miss a target because of the potentially fatal

consequences, such as in the case of IED detection, then the response criterion will be set to a level where misses are rare but false alarms are more likely. The response criterion can be affected by a driver's expectancy that there will be an IED or other threat, such as if the area they are driving through is prone to IED attacks.

IED detection can be modeled using signal detection theory. In this experiment, IEDs and distracters were presented and participants were tasked with identifying whether the item presented was an IED or a non-IED. The four outcomes outlined above are all possible outcomes during IED detection. Although no rewards or punishments were administered during this experiment, in reality there are very strong consequences for failing to properly identify IEDs in the field. In the case of this signal detection task, the penalty for misses is more severe than false alarms. The IED performance from this study is likely tempered because the participants were identifying targets in a simulator. In a warzone where the stakes are much higher, the likelihood of false alarms is even greater due to the very severe and potentially fatal consequences of missing an IED in the field. Although there is no way to mimic the severity of an IED in a simulated environment, researchers did their best to imitate the responses to IEDs as much as possible.

Mental Workload Assessment Tools Used in this Study

Given the wide variety of physiological and subjective mental workload measures that are accepted and used in different environments, additional work is needed to determine if a relationship exists between physiological and subjective metrics, and the best combination of metrics for use in an operational environment. This thesis used data from a previous project where researchers collected both physiological and subjective metrics of workload to assess the impacts of workload on performance. If the tools selected for use in the simulator provided robust and statistically significant results, the intention would be to test them in an operational environment. The original data collection effort attempted to avoid primary task intrusion – the extent to which measurement devices interfere with the primary task and contribute to additional workload (Embrey, et. al., 2006).

The subjective metric selected was the Modified Cooper-Harper scale (MCH), a 10-point unidimensional scale that uses a decision tree to guide individuals to appropriate workload ratings. Individuals answer yes or no to a series of decision points and then select the most appropriate rating based on the description provided. Coincidentally, this tool was initially developed for use in a military environment (Embrey, et. al., 2006). The MCH was originally developed to assess aircraft handling but was modified to estimate workload (Gawron, 2000). This assessment can be completed while performing a task or shortly after, so there is minimal disruption to the participant. The MCH has been shown to be ideal for primarily psychomotor tasks (Cain, 2007), which were relevant and required for the simulated driving task performed in this study. Gawron (2000) noted that the MCH was effective at assessing overall workload, and was sensitive to different types of workload.

There were two limitations to MCH that were noted when this tool was selected. First, the MCH assumes that it is desirable to have a low workload, which is not necessarily a valid assumption. This assumption is manifested in the scale by directing individuals to rate their workload at a “one” if minimal effort is required, “two” if low effort is required, and so on. Recall that the Yerkes-Dodson Law would suggest that an intermediate level of mental workload would lead to optimal task performance. On this scale, the intermediate workload ratings (four to six) describe workload as between moderately high and maximum in order to still maintain performance (Gawron, 2000). However, the optimum level of workload varies widely depending on the specific task. Also, the MCH has no diagnostic ability which would limit its usefulness in field environments or to help as a benchmark when scheduling activities based on workload levels (Embrey, et. al., 2006).

Electroencephalography (EEG) was selected as one physiological metric to assess mental workload. According to Berka, Levendowski, Cvetinovic, Petrovic, Davis, Lumicao, Zivkovic, Popovic, and Olmstead, (2004), EEG is the only physiological signal that accurately reflects subtle shifts in alertness, attention, and workload that can be identified and quantified on a second-by-second time-frame. Although EEG data can provide useful results, collecting good data is often difficult to do. EEG data can be very easily contaminated by a number of artifacts including facial muscle or eye movements.

EEG measures the variation in voltage over time of the brain, recorded through the use of electrodes placed on the scalp. The presentation of varying stimuli will elicit a voltage change in the EEG reading. The term event-related potential describes these voltage changes as the result of a certain event (Coles and Rugg, 1995) and is another method used to assess an individual's mental workload.

There are many different approaches to analyzing mental workload using EEG. One method is an EEG frequency analysis which would be used to assess an operator's state, such as arousal during a vigilance task. There are four types of waves that are classified based on their frequency. Delta waves are found up to 4 hertz; theta waves from 4 hertz to less than 8 hertz; alpha waves from 8 to less than 13 hertz; and beta waves are anything greater than 13 hertz (de Waard, 1996). Murata (2005) found that alpha and beta are generally elicited from increased alertness during a period of higher mental workload. Examining wave presence and patterns can allow researchers to determine the amount of mental workload induced by certain tasks.

Two different methods exist to elicit event-related potentials. One method is through primary task stimuli, while the other involves an additional and less important secondary auditory task or visual probe. The P300 component has been shown to increase in amplitude with the increase in difficulty of the primary task. However, the P300 component elicited by the secondary task has a decrease in amplitude with the decrease in difficulty of the primary task (Kramer, Trejo, and Humphrey, 1995). The downside to using P300 to assess mental workload is the limited amount of time that information processing is actually conducted in the central nervous system. Therefore, the P300 component is subject to missing changes in workload (Murata, 2005). Kramer, et. al. (1995) found the P300 component to be "sensitive to changes in processing demands." Additionally, P300 is one of the more sensitive components of event-related potentials, as compared to N100 or P200 (Miller, Rietschel, McDonald, and Hatfield, 2011).

A third approach to measuring mental workload with EEG is through the use of wavelet transformation. This technique involves using the Fourier transform to understand how the ratio of a specific band frequency changes with differing levels of mental workload. Wavelet transformation allows the signal to be viewed over a dynamic window of time, allowing for higher resolution and a better ability

to identify abrupt changes in EEG. Wavelet analysis allows for appearance time and total power to be extracted, which are reliable metrics for assessing changes in workload in a variety of human-computer interaction environments (Murata, 2005).

Attempts to use EEG to analyze mental workload in an operational environment have been very limited at this time. Currently, the majority of studies have been conducted in a controlled laboratory environment. The application of these results to the real-world is unclear at this time, primarily due to the fact that operational environments are highly variable (Kohlmorgen, Dornhege, Braun, Blankertz, Müller, Curio, Hagemann, Bruns, Schrauf, and Kincses, 2007). However, one study found significant correlations between EEG indices of mental state changes and performance in labs, simulations, and operational environments (Gevins, Smith, McEvoy, and Yu, 1997; Gevins, Smith, Leong, et al., 1998). This suggests that further studies using EEG to assess mental workload may provide useful results in a variety of environments.

Eye blinks are another physiological eye metric that be used to measure mental workload. These data can be collected either through the use of a camera-based eye tracker or through electrooculogram (EOG). In this experiment, the camera used to record eye metrics collected eye blinks, fixations, and saccades. From these recorded eye metrics, blink frequency and blink duration were the eye metrics used to assess mental workload. The lack of physical control over eye blinks is one reason they are selected as a metric to assess mental workload. Blink frequency has not been used as extensively to assess workload; however, there is evidence suggesting that blink frequency is affected by detection and identification tasks. Several studies involving eye blinks have shown that as mental workload increases, blink frequency increases and blink duration decreases. A study using a mental arithmetic test found significant differences in blink frequency as the difficulty of the task increased (Chen, Epps, and Chen, 2011). Variations in blink frequency and duration in surgeons have been correlated with mental fatigue, lapsed attention, and stress overload (Zheng, Jiang, Tien, Meneghetti, Panton, and Atkins, 2012). Benedetto, Pedrotti, Minin, Baccino, Re, and Montanari (2011) proved blink duration to be a sensitive and reliable indicator of visual workload as well as a diagnostic tool to assess driver workload.

Purpose of Study

There were three main objectives of this study. The first was to identify if mental workload was one of the major contributors to poor mission performance, such as the inability to detect improvised explosive devices (IEDs). Researchers were interested in determining if the requirements of MRAP vehicle drivers exceed their capabilities to do their job and keep fellow Warfighters safe. Confounding factors related to workload are driver training received, driving environment, and driving demands such as communication within one's MRAP and between vehicles. The second objective was to determine the ability of heavy vehicle drivers to appropriately assess their workload level by measuring the statistical significance of the results from physiological measure, performance measures, and subjective workload measures. Physiological measures could be correlated to assess validity and interchangeability. Also, if physiological and subjective measures varied as expected, it was possible that a quick, subjective tool could be used to immediately assess an MRAP vehicle driver's mental workload to a reasonable level of accuracy and sensitivity. The third objective was to identify if heavy vehicle drivers' mental workload and performance varied significantly based on their level of expertise. Researchers expected more experienced drivers to be able to perform under higher levels of workload than novices. The differences in physiological workload, subjective workload, and performance have not been quantified.

Statement of Hypotheses

It is hypothesized that:

- If experimental conditions under which blink frequency and electroencephalography measures of mental workload are constant, then these two measures of mental workload will be correlated.
- If significant differences exist for average blink frequency between the four course segments, then significant differences will also exist between subjective workload ratings for the four segments.
- If simulated events are consistent across all participants, then participants who are novice drivers will experience higher levels of perceived workload than expert drivers.

- If mental workload is increased, then the miss rate and false alarm rate for IED detection will increase.

Methods

Participants

The participants for this study were Marine Corps personnel who ranged in rank from Marine Lance Corporal through Staff Sergeants (E3-6) with a minimum of two combat deployments to Operation Enduring Freedom (OEF) or Operation Iraqi Freedom (OIF). All participants had heavy vehicle driver experience in an MRAP, Stryker, Abrams, Bradley or Heavy Equipment Transporter (HET) during at least one deployment. All volunteers were licensed and qualified drivers. Participants were U.S. citizens between the ages of 20 and 30. They had normal or corrected to normal vision and normal hearing. Fourteen military volunteers participated in this study (N = 14).

All participants were recruited on a voluntary basis from Camp Lejeune, North Carolina. There was no compensation for participation nor were individuals penalized for not participating. All participants signed informed consent documents in order to demonstrate their agreement to participate in the study. Individual identification numbers were used to protect the identity of the participants.

Materials

This mental workload study utilized subjective and physiological (objective) metrics to collect Warfighter workload data while using a Ride Motion Simulator (RMS). Subjective data were collected via the Modified Cooper Harper (MCH) scale that participants filled out to monitor their perceived workload level after specific events occurred throughout the simulator drive. Physiological data were collected from electroencephalography (EEG) and eye trackers were used to eye characteristics, such as blink rate, blink duration, saccade length, saccade duration and fixation duration. The physiological metrics were continuously recorded throughout the entire experiment.

Video data were captured using video cameras mounted on the inside cabin of the simulator. Audio data were also recorded in order to capture responses and communications from the participant. Participant responses were collected using a steering wheel and microphone.

Ride Motion Simulator

The Ride Motion Simulator (RMS) is a six degree-of-freedom motion-based simulator capable of reproducing the dynamics of military ground vehicles over a vast array of terrains (Figure 1) seen by current force vehicles. Located at the Tank and Automotive Research, Development and Engineering Center (TARDEC) in Warren, Michigan, the simulator is comprised of a platform mounted on a hexapod design base producing longitudinal, lateral, vertical, roll, pitch, and yaw motion. The simulator cab is re-configurable and has the capability to collect performance data. The RMS has been safety certified to permit use by soldiers and experimenters.

Figure 1: Ride Motion Simulator (RMS)

The simulation environment was constructed to present participants with visual, motion and audio cuing to recreate a realistic driving experience. The simulator cab was configured with a Surrogate Common Crew Station which includes a vehicle seat, a seat belt, a yoke, and flat panel displays. The temperature was typically normal room temperature. Audio cuing was limited to presenting the participant with the commander's voice and the vehicle's sounds (engine noise correlated to engine revolutions per minute (RPM)). A military standard (MIL-STD-1474D (1997), "Department of Defense Design Criteria Standard – Noise Limits") was used as a guideline regarding noise exposure and hearing protection requirements.

The driving scenario consisted of driving through a simulated environment which consisted of four segments. The driving course between two checkpoints was classified as one segment. The order of segment presentation was the same across all participants. The difficulty of the terrain and frequency of communications and IED placements increased from segment one to segment four. Participants experienced relatively easy driving conditions at the start of the experiment, concluding with a full ambush by the end of the experiment. The increase in the frequency of events and difficulty of the driving environment was intended to increase the participant's workload.

Modified Cooper Harper (MCH) Scale

Subjective mental workload data were collected using the MCH scale. An example of the scale that was used in this experiment can be found in Appendix C. The MCH scale uses a decision tree to record an individual's perception of the workload associated with specific, preselected tasks or series of events. For this experiment, in order to minimize interference with the scenario, the scale shown in Appendix C was administered verbally after each of the four segments during the experiment. The MCH was administered at the latter checkpoint of each segment and was intended to reflect the workload experienced between that checkpoint and the previous one of each segment.

Electroencephalography (EEG)

All of the participants wore an EEG BioSemi electrode cap system with pre-amplified surface electrodes sampling at a rate of 500 hertz (Hz) throughout the simulation event. The EEG data were acquired using a 24-bit, 72-channel ActiveTwo amplifier and ActiView software. A picture of the device is shown below in Figure 2. EEG recording sites were prepared in accordance with the standardized international 10-20 electrode placement system (Nuwer, Aminoff, Desmedt, Eisen, Goodin, Matsuoka, Mauguière, Shibasaki, Sutherling, and Vibert, 1994). A water-soluble, salinated electrode gel was inserted into each of the electrode casings to facilitate conductivity between the scalp and electrode surfaces. Vertical electrooculogram (VEOG) and horizontal electrooculogram (HEOG) eye movements were monitored using bipolar electrode montages attached superior and inferior to the right eye (VEOG) and

both orbital fossa (HEOG). These eye movements were recorded to determine which specific aspects of the EEG recordings were directly attributed to eye movements.

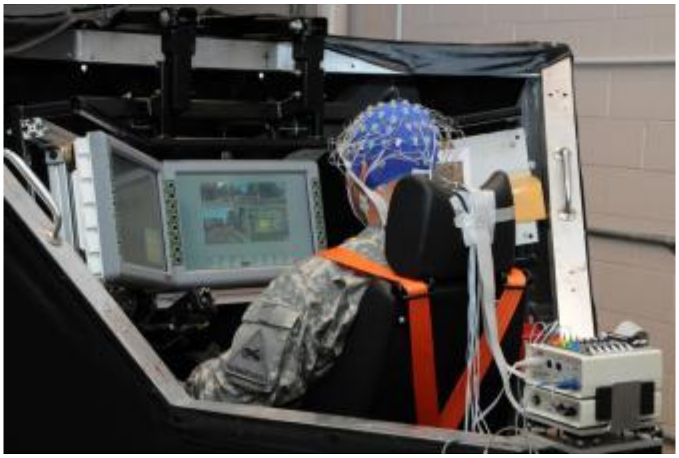


Figure 2: EEG BioSemi electrode cap system

Blink Frequency/Duration

Participants' blinks, fixations, and saccades were recorded via a camera-based, non-contact tracking system shown in Figure 3. This system allows for observing the natural participant eye and head movement behavior at adequate spatial resolution (approximately 0.5°). Eye and head movements and measurement reliability data were recorded in real time and synchronized with the other data measures.

Figure 3: TARDEC Eye Tracker

Tasks and Stimuli

The simulation environment was constructed to present participants with visual, motion, and audio cuing to recreate a realistic driving experience. Volunteers for this experiment served as the role of a vehicle driver. They were seated at a crew station, and were able to control the direction and speed of the vehicle as it navigated through a simulated environment. The experiment simulated a Stryker, or light-skinned combat vehicle, traversing cross-country terrain with maximum acceleration not exceeding 2 g's (g = acceleration due to gravity, 9.8 m/s^2). The motions experienced by the test volunteers did not exceed ranges beyond ± 20 inches in the translational directions and ± 20 degrees. The simulator's safety interlock system was set to ensure that the ride motion did not exceed these position or acceleration levels. The driver was equipped with a headset that allowed audio communication to be presented to them in order to simulate the dynamic communications among members of the battalion. Experimenters were also able to communicate through the audio communication system, as well as maintain sight of the participants via camera views and direct vision, at all times throughout the experiment.

Design

Participants performed a primary task with secondary tasks. The primary task was driving and the secondary tasks were communications and IED detection. These tasks were all consistent with M-ATV driver tasks where vehicle operation is the primary task and the secondary tasks are monitoring communications and making return calls, in addition to monitoring and avoiding known, suspected, and unknown IED emplacements, as workload permits.

The experimental design was 2 x 2 within-subjects design. The within-subjects approach was chosen because of the limited number of experienced operators available to support this research. The independent variables were communications activity (Comms) and IED detection with high and low levels for each. As the simulation progressed, the frequency of communications and IEDs increased.

Table 2 classifies how the segments increased in difficulty from segment one to segment four. The table indicates the number of targets and distracters presented in each segment per segment as well as the number of statements or dialogs to which the participant was expected to respond. The first segment

was intended to present drivers with low communications activity and low IED presence. The second segment had higher communications activity but still a lower IED presence. The third segment consisted of lower communications activity and a higher IED presence. The final segment was intended to provide higher communications activity and higher IED presence. There were three checkpoints that defined the four segments. Targets included threats such as IEDs and hostile individuals. Distracters were non-threats such as trash and civilians. The segments increased in difficulty because the number of targets and their dialog responses increased from segment one to segment four. The number of distracters was variable during the four segments, but was the most intense in the second and third segments. Segments two through four each concluded with an ambush. As the segments progressed, the targets presented became more frequent and also more dangerous.

Segment	Targets	Distracters	Dialog Responses
1	0	5	4
2	7	12	11
3	7	10	17
4	8	6	35

Table 2: Classification of Segment Difficult

For the overall study, many variables were collected by the simulator, the EEG equipment, the eye tracker, and the researchers. For the purpose of this thesis, five specific dependent variables were collected from the simulator experiment. The first dependent variable was EEG workload which was derived from the raw EEG data. Two additional dependent variables were blink frequency and blink duration, measured by the eye tracker. Fourth was the Modified Cooper Harper scale (MCH) rating obtained verbally from the participants during the experiment. Finally, the IED miss rate and false alarm rate were derived from participant responses during the experiment. The IED detection performance was used to analyze the participants’ performance.

Procedure

Participants traveled to the Tank and Automotive Research, Development and Engineering Center (TARDEC) in Warren, Michigan and reported for an experimental session that was expected to last

approximately five hours. During the session, participants were briefed about the experimental procedures and received a safety briefing on the Ride Motion Simulator. The participant was then given the opportunity to read the consent form, and if they agreed to participate, they signed the consent agreement. Upon giving consent, the participant questionnaire was administered to ensure compliance with the experiment selection criteria found in Appendix D. This process lasted 15 to 20 minutes.

Upon consenting to the experiment, participants completed a safety brief, then were prepped and fitted with the EEG electrode cap. Participants were encouraged to ask any questions or alert the experimenter if any discomfort occurred and needed to be corrected. The configuration, or calibration, script for the eye-tracking system was also completed. After all equipment was set up, participants completed a practice run on the Ride Motion Simulator. Total setup time duration lasted 35 to 50 minutes.

Participants sat before the driver crew station monitors. They were able to communicate with the experimenters via headsets connected to the Audio Recording System. The participant completed several missions through the simulated environment. Participants on the simulator were verbally asked to rate their workload on the MCH scale after each segment was completed. Experimental testing lasted no more than three hours. At the end of the testing, participants were encouraged to ask any questions about the experiment as well as provide any feedback about the experimental session.

Originally, the scenario was intended to have automated communications that came on at predetermined intervals in response to certain events. The scenario for the actual experiment was read by confederates that assisted with the experiment. This was due to the fact that some of the automated communications may have questioned participants about targets or distracters that they did not identify. Confederates represented different vehicles in the convoy and read their script or responded to participant communications as appropriate. Consequently, it was possible that the communications were not consistent across all participants.

With the scenario modification to be read by confederates, there was a potential confound because the script reading varied between individual confederates. Specifically, consistent communications could not be maintained across all participants. There may have more or less

communication depending on how much the participant deviated from the script by asking questions, initiating conversation or getting wrapped up in other tasks. This variability could not be controlled for because the participants were interacting with actual people. Additionally, depending on the confederate's level of military experience, different jargon may have been used or differing senses of urgency communicated to the participant. Given the variability between confederates and that the confederates were not consistent across all participants, this could lead to potential variability in participant workload.

Another potential confound was that the volume was not consistent across all testing days. On days when the simulator volume was very loud, it was clear that the insurgents were firing in the participant's direction. However, on certain days when the simulator volume was very low, there was a more significant delay in participants being able to visually identify that the insurgents were firing at them. This would have affected their reaction time in response to the event. Additionally, louder noises may have the effect of elevating stress level. As mentioned previously, stress can be a confounding factor related to workload.

Data Analysis

EEG Analysis

The EEG data were analyzed with EEGLAB, an open source toolbox that runs through MATLAB. EEGLAB supports BioSemi data, so the raw EEG data were imported directly into EEGLAB as it was collected from the simulator.

First, a channel map was created using EEGLAB in order to determine the exact positioning of the electrodes on the participants' heads. This channel map was used to select the appropriate electrodes for the time-frequency plots and is shown below in Figure 4. The frontal midline has been shown to reflect changes in workload; therefore, channels 4, 11, 38, 39, 46, and 47 were selected for this analysis. Additionally, two frequency waves—alpha and beta—were analyzed on each channel. Alpha waves change with visual action and cognitive activity and changes in beta waves are a reflection of alertness and cognitive activity (Berka, Levendowski, Lumicao, Yau, Davis, Zivkovic, Olmstead, Tremoulet, and Craven, 2007; Onton, Delorme, and Makeig, 2005).

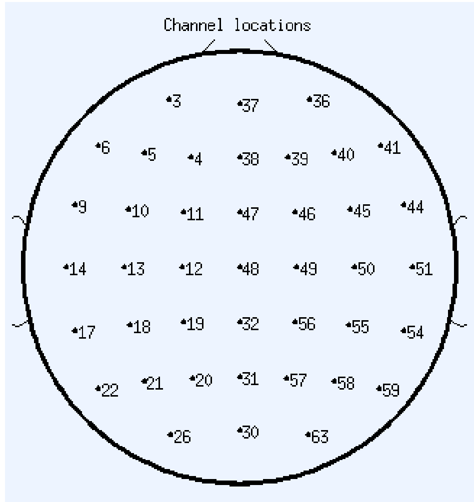


Figure 4: EEG Channel Locations

Next, a channel time-frequency plot was created on the preselected channels to allow the data to be visually analyzed for changes in brain activity. This allowed event-related spectral perturbation (ERSP) to be used. ERSP was used to study changes in EEG activity surrounding events. It reflected how the power at different frequencies in a signal was altered in relation to a specific time point, such as a signal. Positive values showed an increase in power while negative values showed a decrease in power relative to the baseline (Makeig, 1993).This analysis process was performed for each individual subject across all channels and frequencies.

Signal Detection Analysis

The events that were presented during the driving task were classified in order to be used for signal detection theory. If a target was presented on the screen, there were two possible responses from the participant. If they identified the target as such, this was classified as a hit. If the participant failed to respond to the target, this was classified as a miss. When distracters were presented on the screen, there were also two possible responses. If the participant identified the item as a distracter, this was classified as a correct rejection. If the participant identified the item as an IED, this was classified as a false alarm. If participants failed to respond to distracters, these data were not classified as a correct rejection, but were also not counted against their performance.

Signal detection theory was used to calculate sensitivities and response criteria for each participant. Participant sensitivity assessed how easy it was for a participant to identify an IED as such when faced with distracters and other competing tasks. The response criterion determined how much evidence was needed for a participant to identify an item as an IED or a distracter.

Analysis of Variance

To determine any overall significant relationships within the physiological and subjective metrics, a 2 x 4 mixed-design analysis of variance (ANOVA) was run. The between-subjects variable was experience and had two levels—novice and expert. Novice drivers had approximately five years less heavy vehicle driving experience than expert drivers. The two categories of driver experience were tested to confirm that they were significantly different from one another. The workload segment was the within-subject variable and the four levels were the four segments as previously shown in Table 2. This ANOVA design was run for five dependent variables—subjective workload rating, IED miss rate, IED false alarm rate, average blink frequency, and average blink duration. The IED miss rate was the percentage of unidentified IEDs out of the total number of IEDs presented during the trial, and the IED false alarm rate was the percentage of total IEDs that were not actual IEDs but identified as such. It was not necessary to evaluate hits and correct rejections because these results are complementary to misses and false alarms, respectively. From these ANOVAs, the interaction effects were examined to determine overall significant differences between segments and experience levels. After any significant findings were determined, Tukey Tests were conducted to determine which levels of the factors were significantly different.

Eye Blink Analysis

The eye blink frequencies and duration were also analyzed using two-sample t-tests. The results were grouped by “yes” responses—hits and false alarms—and “no” responses—misses and correct rejections. The two-sample t-tests were used to compare the results from the “yes” responses to the “no” responses. Responses were collapsed across all four segments for a larger sample size since the quality of the eye data in smaller quantities were questionable. The intention was to determine if there are

differences in the eye's response when individuals believe they identify something as a threat versus an item that is not a threat.

The amount of ***missing*** eye tracker data were calculated for each participant. The eye tracker measured three different eye metrics—blinks, fixations, and saccades. At any given time, the eye will be in one of these states, so a value will be recorded for one of these states at all times. Missing values in all three categories would indicate that the eye tracker was not recording the participant's eye information. The table below shows the percent of time when the eye tracker could not record the participant's eye information. The large amount of missing data impacted the results that were concluded regarding eye metrics.

Participant Number	**% of Data *Missing***
1	82%
2	87%
3	74%
4	65%
5	99%
6	99%
7	10%
8	99%
9	76%
10	92%
11	62%
12	76%
13	91%
14	89%

Table 3: Percent of Missing Eye Tracking Data for Each Participant

As shown in Table 3, there was a significant amount of missing eye tracker data for several of the participants. A serious limitation to the technology used to capture the eye blink information for this study was that the eye tracker was mounted to the computer monitor. There was a very limited range of motion where participants could move their head. Additionally, obstructions such as a participant touching their face or adjusting their glasses obscured the camera from view of the eyes and did not allow any eye blink data to be collected. Other factors contributed to the data loss including system failures and participant

seat height. More recent eye tracking technology uses head mounted cameras, often in the form of glasses, to overcome many of the technical challenges faced with surface mounted cameras. For this study, using a surface mounted camera resulted in a significant amount of lost data. With such large amounts of missing data, it is difficult to achieve statistical significance in the results.

Correlations

The data were intended to be analyzed to determine if a correlation exists between participants' EEG workload measurements and blink frequency. This method would have been used to identify any trends without establishing causation. The EEG workload measurements and blink frequency were both measured continuously throughout the experiment. Unfortunately, due to the qualitative EEG analysis performed, these data were not able to be tested for correlation.

Linear Regression Analysis

Lastly, to determine the predictive power of the selected metrics, a Linear Regression Analysis (LRA) was performed in Minitab. The goal was to individually and collectively use EEG workload, blink frequency and MCH scale ratings to predict the entire sample's IED miss rate performance. Since EEG workload was a more qualitative metric, blink duration was substituted for EEG workload in the LRA as a possible explanatory variable for changes in the miss rate. It was important to identify how workload impacted performance and if certain metrics could be used to predict any performance decrements.

Results

For each segment, participants were exposed to a series of targets and distracters. Figure 5 below shows the targets presented and sample responses in each of the four segments. The bars represent the targets presented and the x's indicate when participant one identified an IED on the course. If the participant has an "x" when there is no target presented, that represents a false alarm. If the participant does not identify a target when it is presented, that is a miss. Each participant drove though the course in a different amount of time, but the targets presented were the same for each participant.

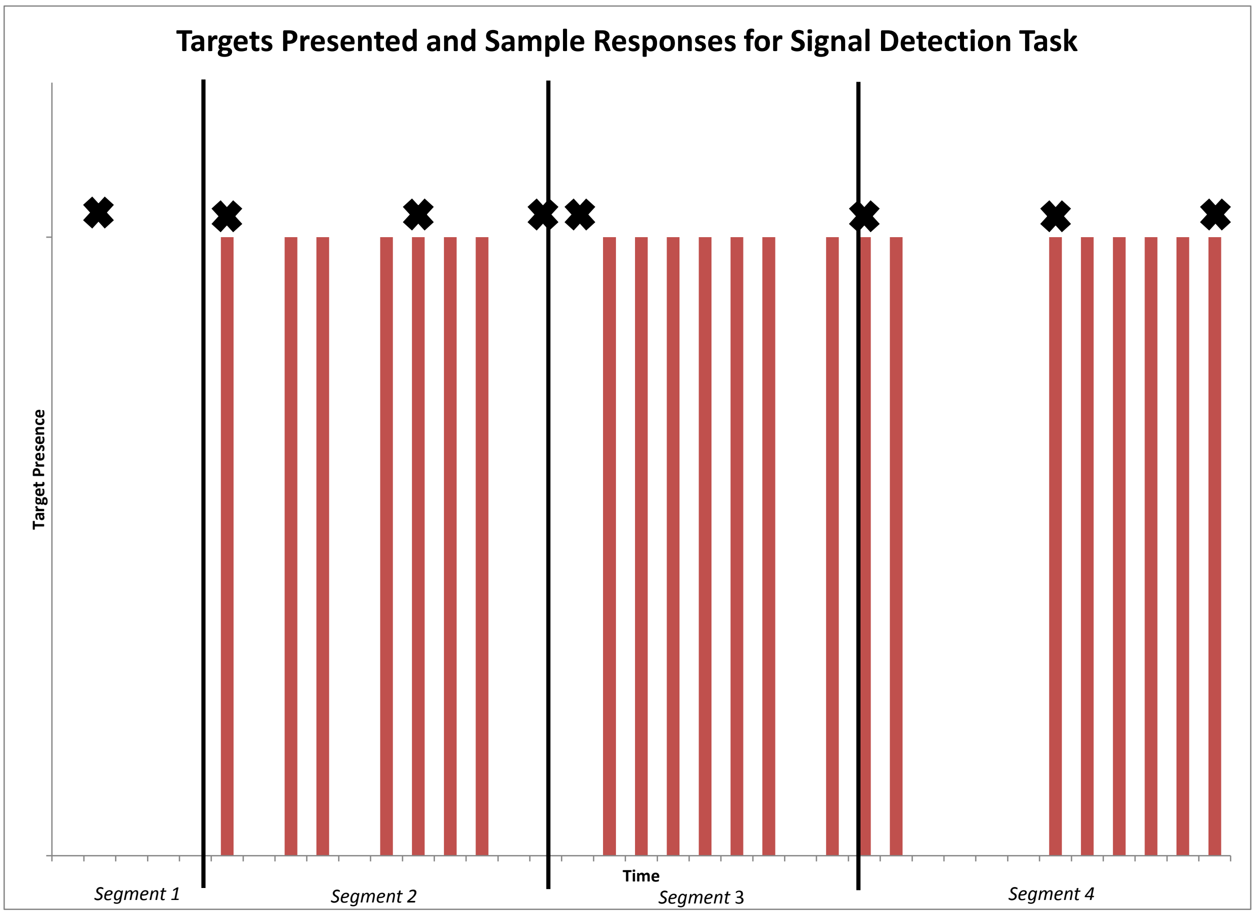


Figure 5: Targets Presented and Sample Responses for Signal Detection Task

A series of ANOVAs were performed to analyze the significance of several dependent variables. The two factors for analysis were the four segments that were encountered while driving and driver experience level--either novice or experienced. Each variable and any significant findings are explained in detail in the following sections.

Driver Experience

The military experience of the 14 participants was graphed in order to determine who to classify as novice drivers and experienced drivers. There is currently no formal classification method to distinguish novice drivers from experienced ones. Figure 6 of subjects' military service showed a separation at six years of service. The subjects were separated into two groups based on this distribution. It was determined that participants with less than six years of experience would be classified as novice drivers. Those with more than six years of experience were considered experienced drivers.

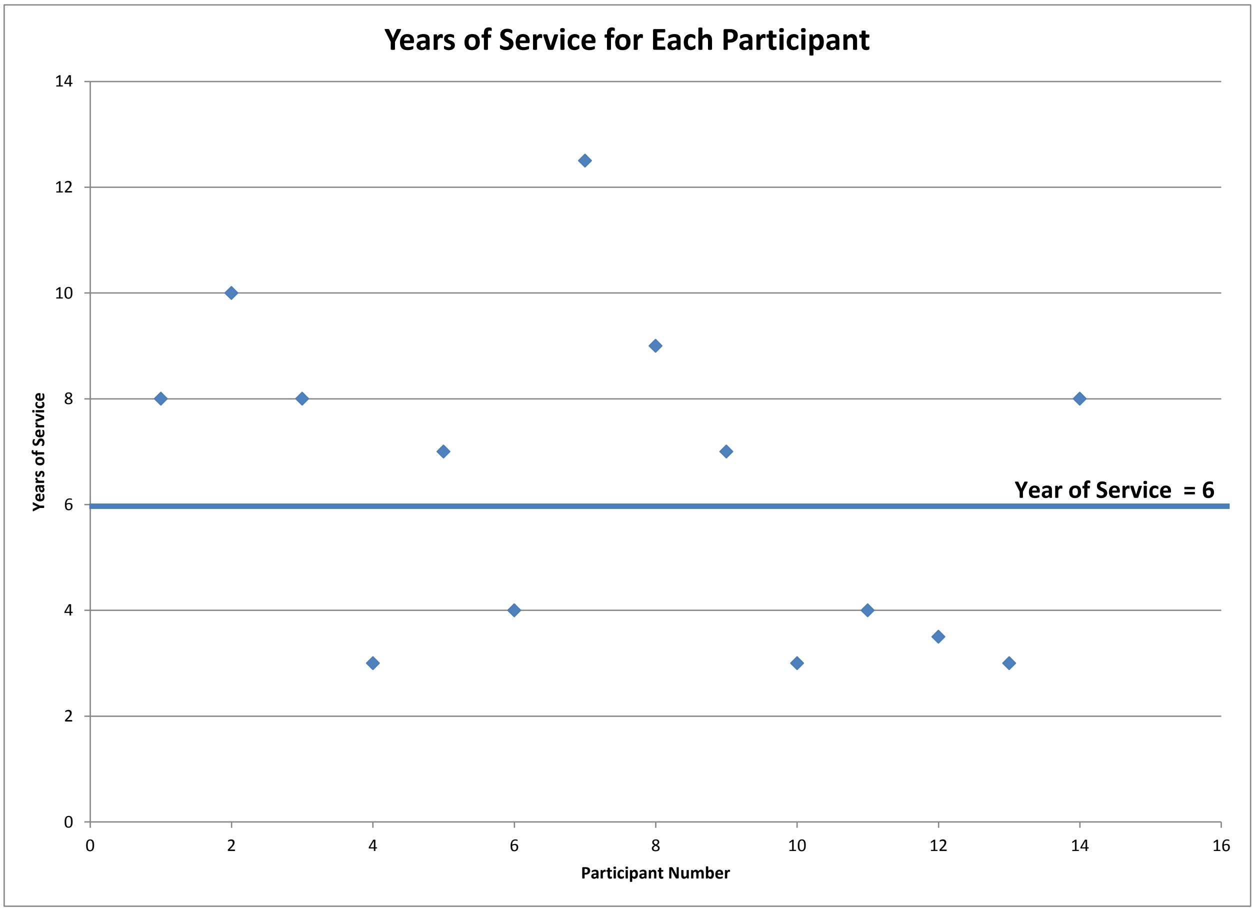


Figure 6: Graph of Years of Military Service for 14 Study Participants

After classifying the two experience levels of the drivers, the mean years of service and standard deviation for the two previously defined groups were calculated. These results are shown below in Table 4.

Level	Mean	St. Dev.
Novice	3.42	0.49
Experienced	8.69	1.83

Table 4: Mean and Standard Deviation of Years of Service

A two-sample t-test was performed in Minitab to determine if these two groups were significantly different based on their years of service as classified by the experience levels. It was concluded that these two groups are significantly different from one another ($t(8) = 7.78$, $p = 0.000$). Figure 7 graphically shows the mean and standard deviation of these two groups.

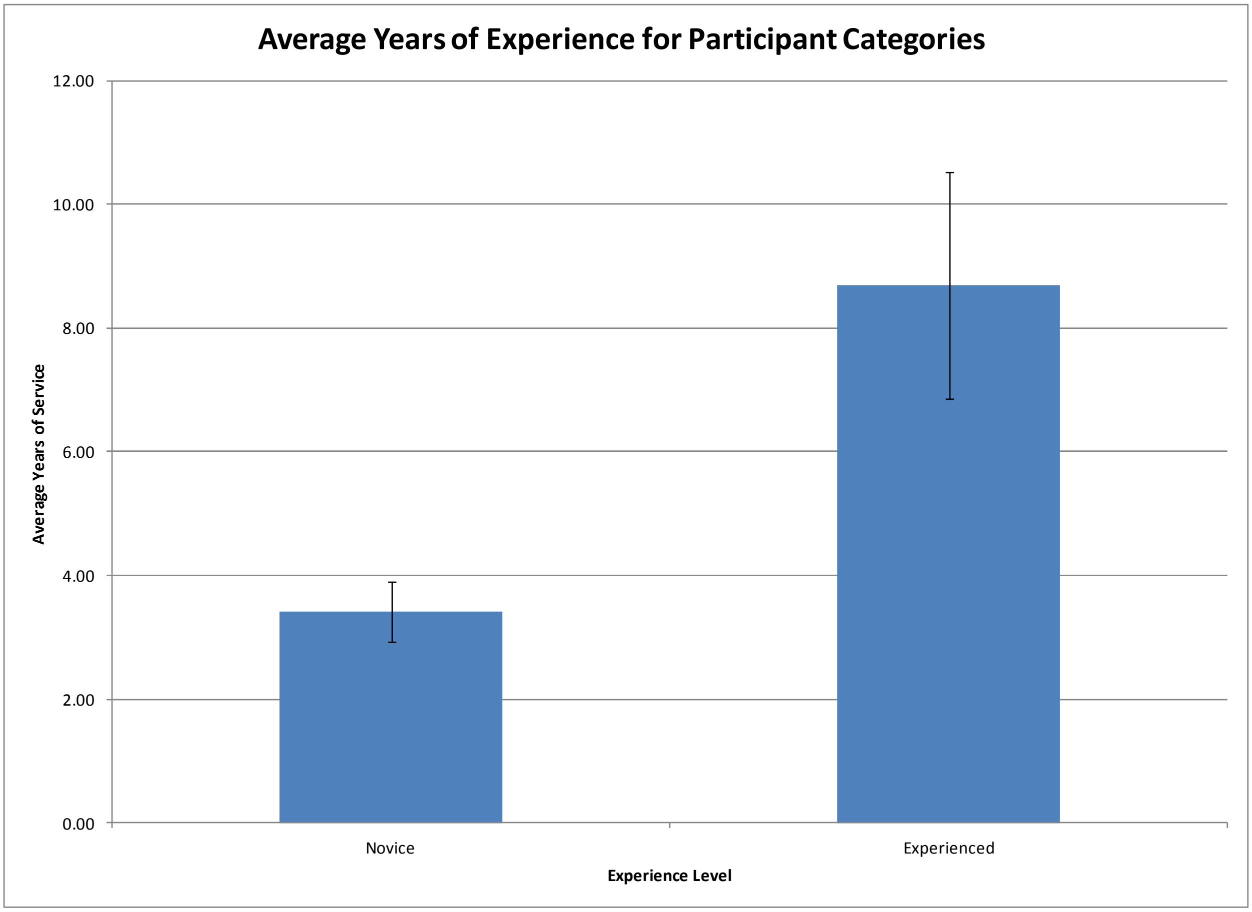


Figure 7: Average Years of Experience for Participant Categories

Modified Cooper Harper Scale Rating

Subjective rating scales, such as the Modified Cooper Harper scale, generally produce ordinal results, meaning only a ranking can be established between numbers. The scale does not indicate how

much more difficult a certain rating is over another. Additionally, one cannot determine the magnitude of difference between any two numbers on the rating scale. Several studies performed by Wierwille and Casali (1983) determined that the Modified Cooper Harper scale was able to discriminate between all 10 load levels. It was found to be a statistically reliable indicator of workload. Wierwille and Casali validated the use of interval analysis techniques for this rating scale.

Each participant was asked to rate their perceived workload using the Modified Cooper Harper scale at the end of each segment. When the end of a segment was reached, a researcher would ask the participant to rate their workload for the section of course they had just driven. An ANOVA was run for the MCH ratings and it was determined that the ratings were significant for both the segments and driver experience ($F(3,30) = 17.08$, $p = 0.000$; $F(1,30) = 52.95$, $p = 0.000$, respectively). The average MCH ratings are shown in Table 5. The full ANOVA can be found in Appendix E.

	Average MCH Rating
Segment 1	3.2
Segment 2	4.3
Segment 3	5.5
Segment 4	6.2

Table 5: Average MCH Rating between Segments

A Tukey test was performed to mathematically determine which individual segments were significantly different from one another. As shown in Figure 8, the ratings in segment one are statistically different from the ratings in segments three and four. This confirms the expected result that the rating of workload increased significantly as the driver progressed through the course because difficulty increased with each successive segment.

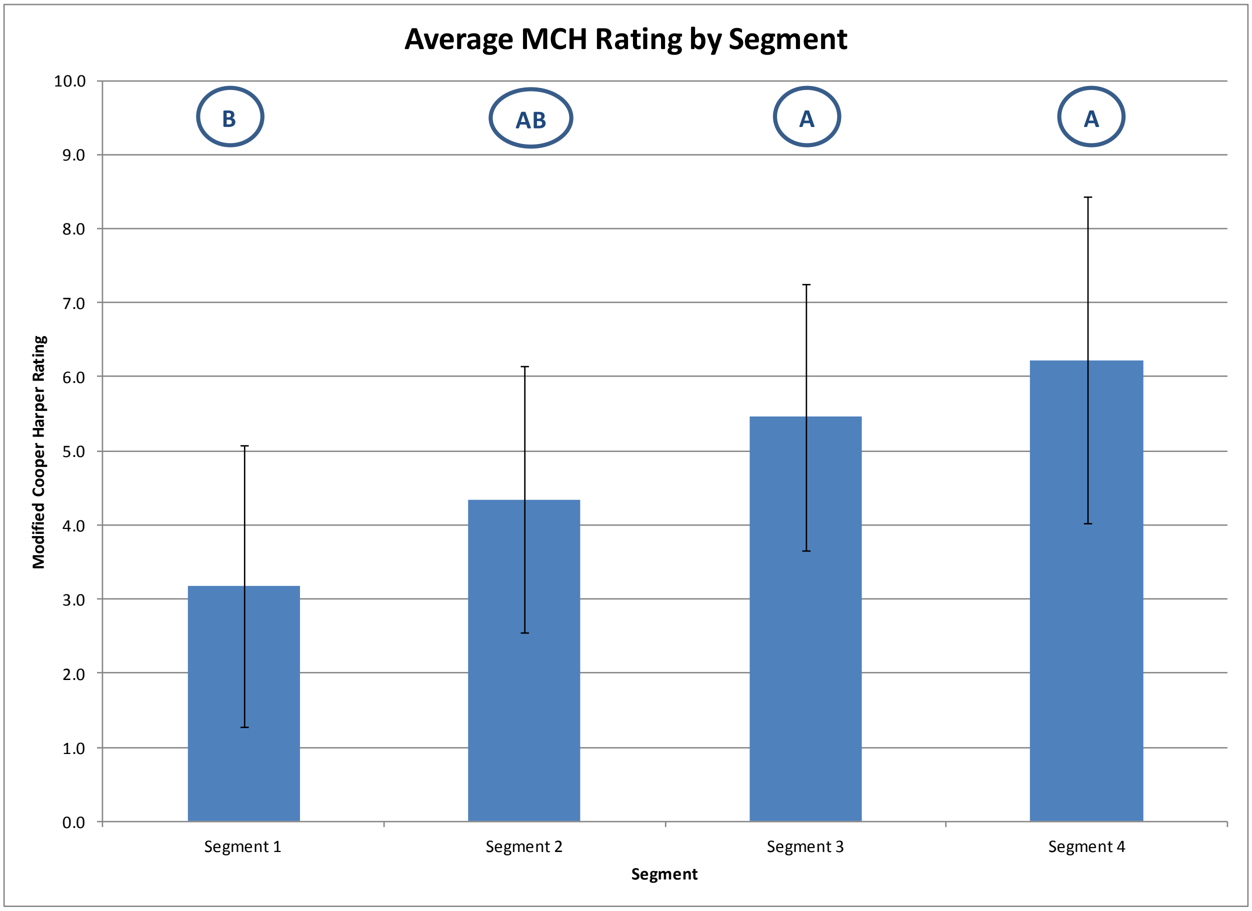


Figure 8: Average MCH Rating by Segment

As previously mentioned, the MCH data grouped by driver experience was also shown to be significant. The average MCH ratings are shown below in Table 6.

	Average MCH Rating
Novice	6.0
Experienced	3.8

Table 6: Average MCH Rating between Experience Levels

A Tukey test was also performed on the two experience levels, which were found to have significantly different MCH ratings. The experienced category of drivers had significantly lower workload ratings than the novice category of drivers. This result was also expected because drivers who are less experienced may perceive more workload because they are not used to as much stress or pressure while driving.

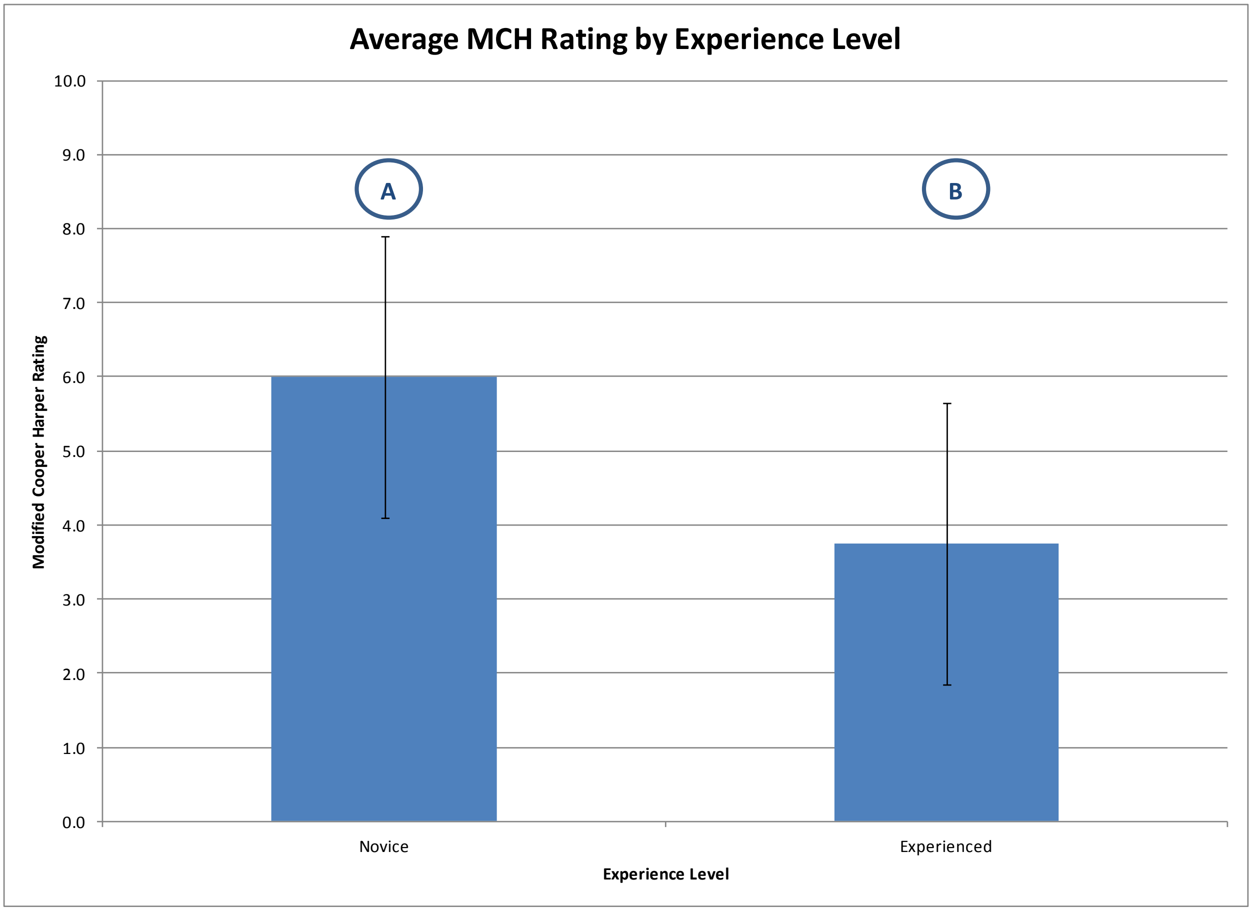


Figure 9: Average MCH Rating by Experience Level

The Modified Cooper Harper ratings were also grouped by segment and experience level. Although the interaction effect between segment and experience level was not significant ($F(3,30) = 0.64$, $p = 0.996$), it was interesting to compare the ratings next to each other. Table 7 outlines the average MCH rating for each segment, grouped by experience level. Figure 10 allows for a direct comparison between the MCH ratings for the two experience levels across the four course segments. It should be noted that both groups rated the segments increasingly more difficult from segment one to segment four. The rate of increase for the novices is slightly greater than the experienced drivers, but not significantly different.

	Average MCH Rating
Novice	
Segment 1	4.3
Segment 2	5.8
Segment 3	6.5
Segment 4	7.7
Experienced	
Segment 1	2.3
Segment 2	3.4
Segment 3	4.7
Segment 4	5.0

Table 7: Average MCH Rating by Experience Level and Segment

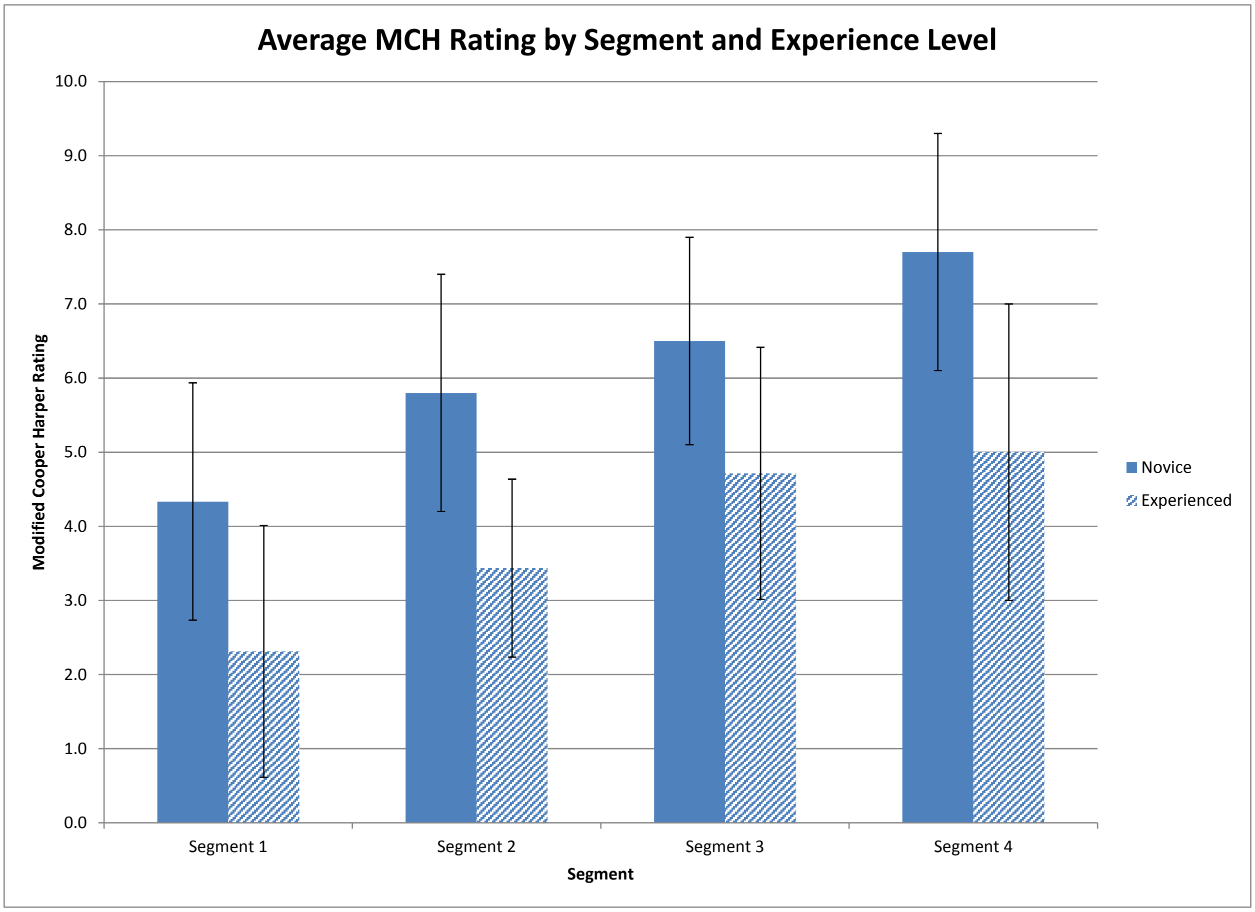


Figure 10: Average MCH Rating by Segment and Experience Level

IED Miss Rate

Driver performance was measured using IED miss rate. A miss was characterized by a target being presented on the simulator screen, but the participant failing to identify it as such. The miss rate was

then determined by dividing the number of misses by the total number of targets for each segment. This rate was determined for each of the four course segments. Similarly, a hit would be characterized by a target being presented on the simulator screen and the participant correctly identifying it. The hit rate can be determined in the same manner as the miss rate. The results for the IED hit rate are complementary to this analysis because the hit rate can be calculated by subtracting the miss rate from one. The individual miss rates were analyzed for trends across each participant, shown in Table 8. Note that the miss rate could not be calculated for segment one because no targets were presented in that segment. An ANOVA was run for the IED miss rate and it was determined that the miss rate was significant between the four segments ($F(3,34) = 28.26$, $p = 0.000$).

		Miss Rate			
Subject	**Experience**	**Segment 1**	**Segment 2**	**Segment 3**	**Segment 4**
1	EXP	--	0.71	1.00	0.63
2	EXP	--	0.29	0.18	0.20
3	EXP	--	0.29	0.43	0.63
4	NOV	--	0.29	0.43	0.67
5	EXP	--	0.57	0.29	0.50
6	NOV	--	0.50	0.43	0.63
7	EXP	--	0.50	0.57	0.63
8	EXP	--	0.38	0.18	0.50
9	EXP	--	0.25	0.38	0.63
10	NOV	--	0.43	0.43	0.33
11	NOV	--	0.25	0.29	0.25
12	NOV	--	0.00	--	--
13	NOV	--	0.29	0.20	0.50
14	EXP	--	0.57	0.14	0.38

Table 8: Individual Miss Rates by Participant

	Average IED Miss Rate
Segment 1	--
Segment 2	0.38
Segment 3	0.38
Segment 4	0.50

Table 9: Average Miss Rate between Segments

A Tukey test was performed to mathematically determine which individual segments were significantly different from one another. As shown below in Figure 11, the miss rate in segment one is statistically different from the ratings in the other three segments. Again, it should be noted that there were no IEDs to be identified in the first segment; therefore, a miss was not possible. Segments two through four are not significantly different from one another. Therefore, the miss rates were not significantly different across the course segments.

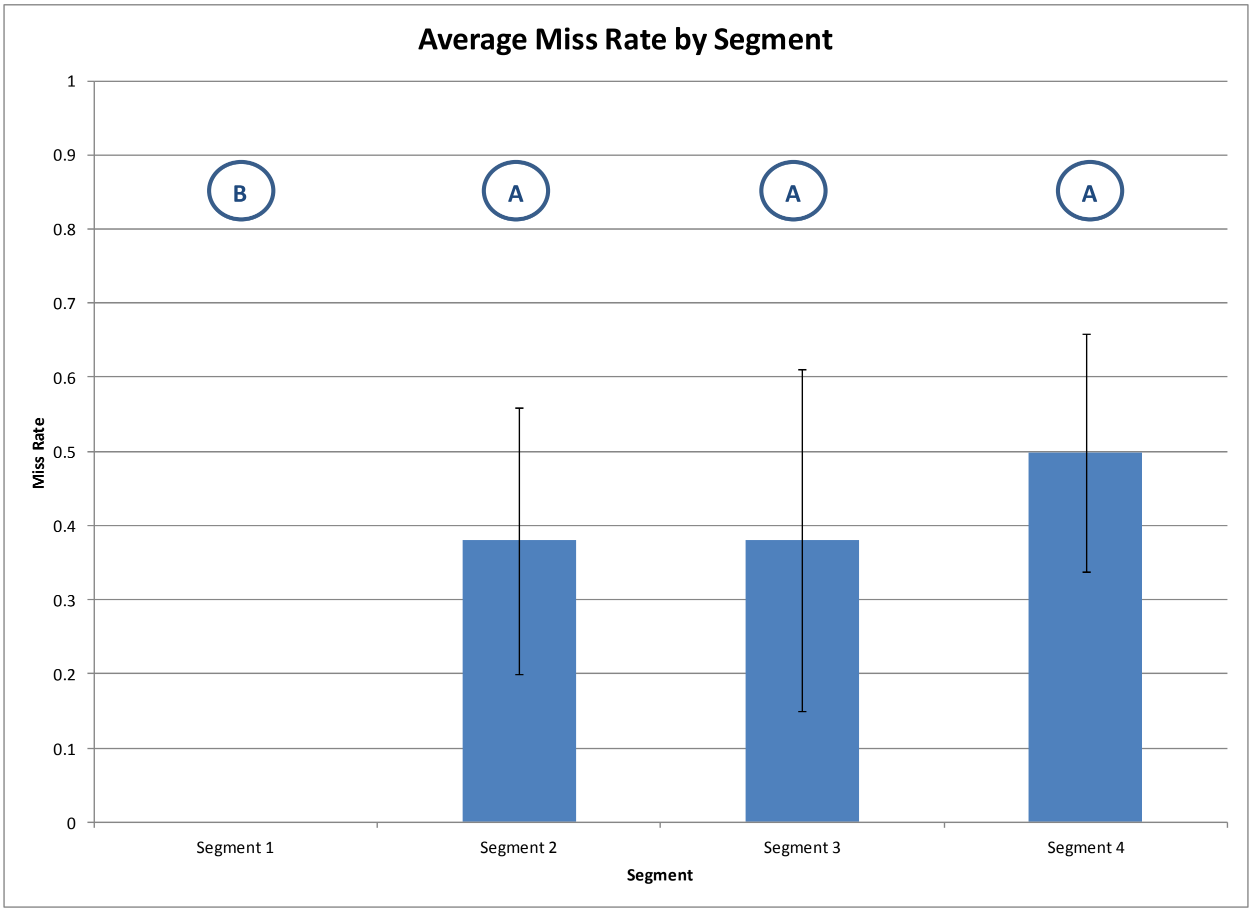


Figure 11: Average Miss Rate by Segment

Segment one was removed from the miss rate analysis. Another ANOVA and Tukey test were run with only segments two through four. The full ANOVA, provided in Appendix E, revealed that the miss rate is significantly different by experience level ($F(1,22) = 5.35$, $p = 0.030$). The experienced drivers actually exhibited a higher miss rate than the novice drivers, shown in Figure 12.

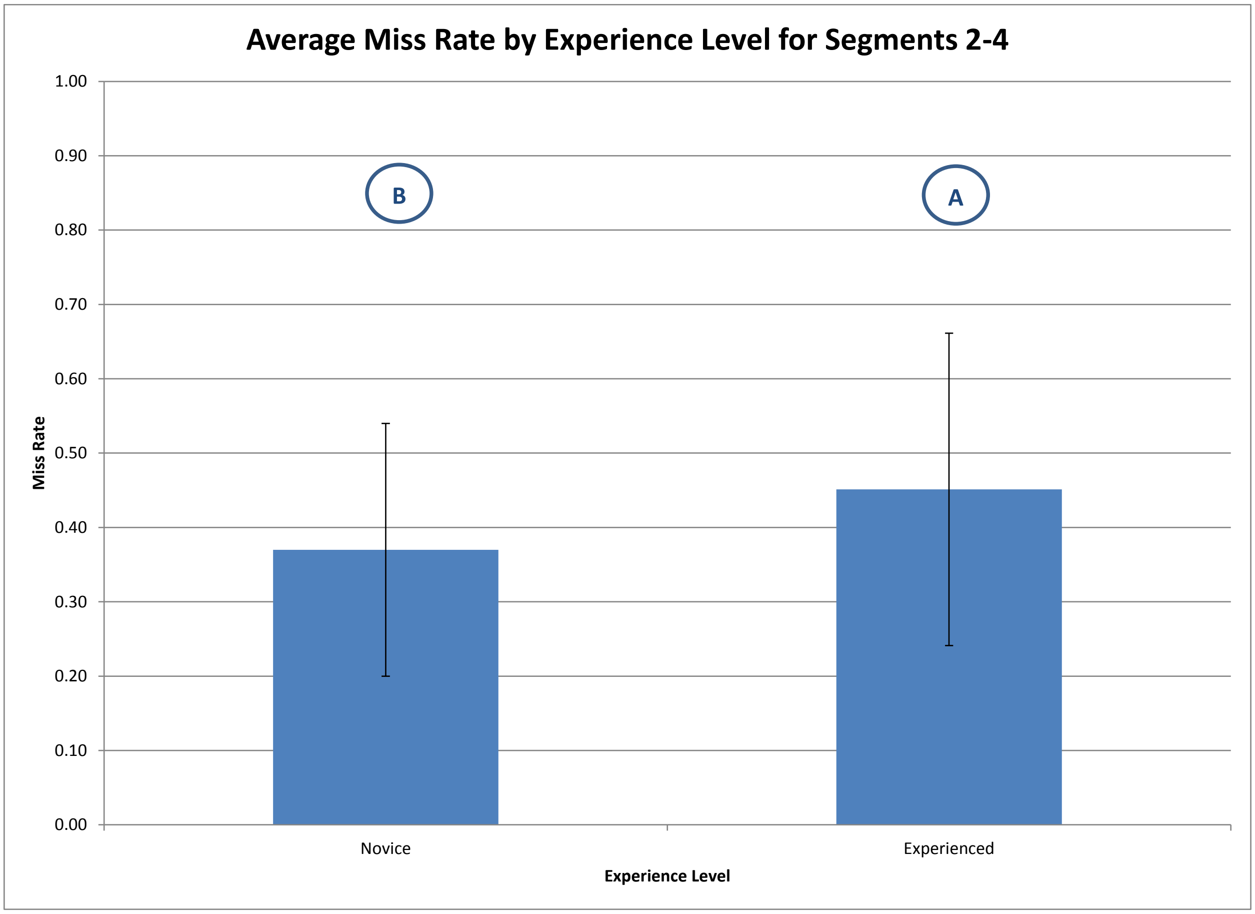


Figure 12: Average IED Miss Rate by Experience Level for Segments 2-4

IED False Alarm Rate

Driver performance was also assessed through the IED false alarm rate. A false alarm was characterized by a distracter or non-IED being presented on the simulator screen, but the participant identifying it as an IED. The false alarm rate was then determined by dividing the number of false alarms by the total number of distractor for each segment. This rate was determined for each of the four course segments. An ANOVA was run for the IED false alarm rate and it was determined that the false alarm rate was significant between experience levels ($F(1,34) = 6.04$, $p = 0.019$).

	Average IED False Alarm Rate
Novice	0.27
Experienced	0.18

Table 10: Average False Alarm Rate between Experience Levels

A Tukey test was performed to evaluate significant differences between the two experience levels. As shown in Figure 13 below, the false alarm rate for the experienced drivers was statistically lower than the false alarm rate for the novice drivers.

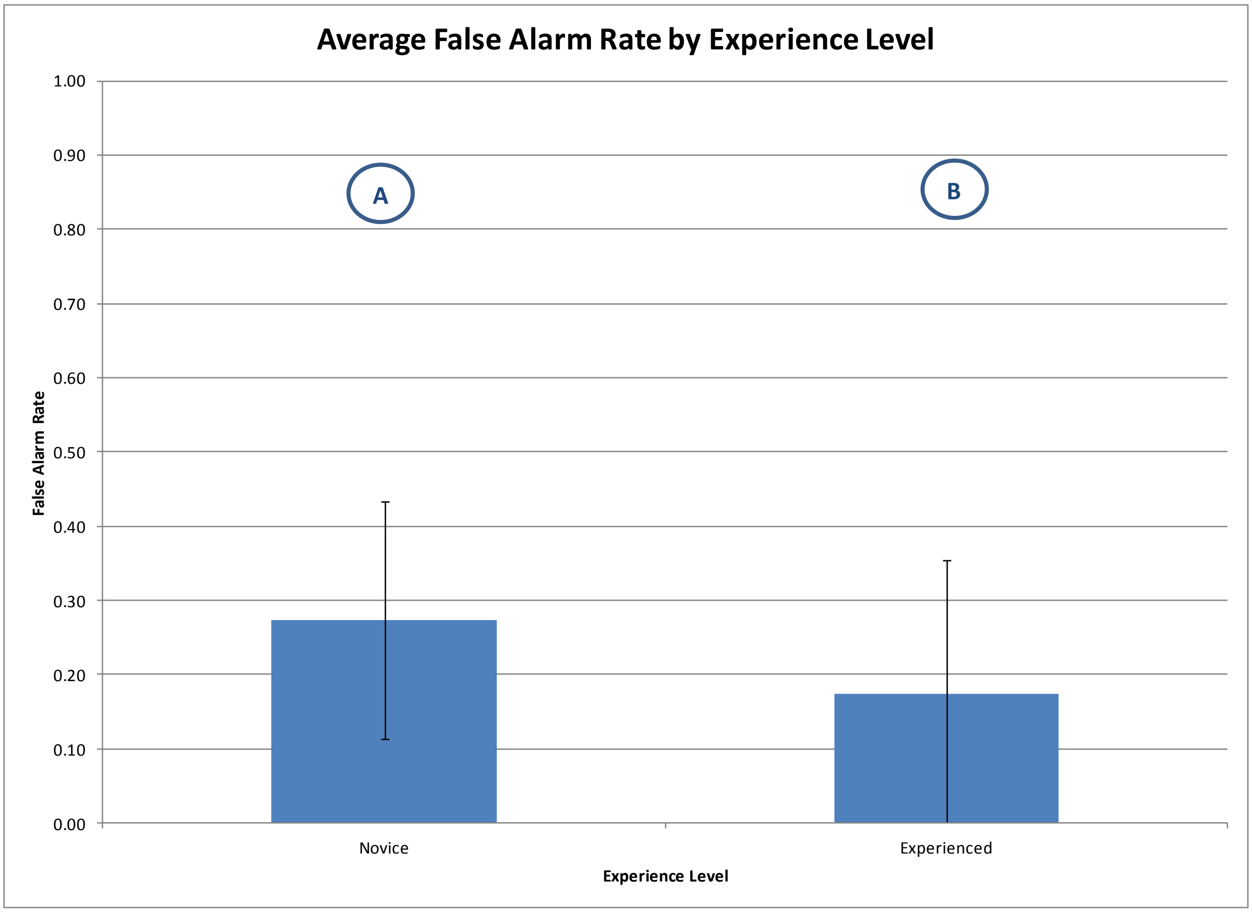


Figure 13: Average False Alarm Rate by Experience Level

Signal Detection Theory Sensitivity

The sensitivity, or d', for each driver was calculated to analyze their IED detection performance. Sensitivity was used to assess how well each participant could distinguish an actual IED from a non-IED. The more correct responses a participant gives, the higher their sensitivity. Equation 1 below shows the formula used to calculate the sensitivity for each participant.

$$d' = Z\big(P(H)\big) + Z(P(1 - FA))$$

Equation 1: Formula to Calculate Sensitivity

As shown in Table 11, the average sensitivity across all participants for each segment decreased as they drove through the course. There were no targets in segment one; therefore, it was not possible to calculate the z-score for the probability of a hit. Consequently, the sensitivity for segment one was not relevant. Table 12 provides the average sensitivity between the two experience levels of the drivers. When broken down by segment in Table 13, the experienced drivers had the hardest time differentiating IEDs in segment three, while the novice drivers had the hardest time distinguishing IEDs in segment four. The sensitivity for the novice drivers declined, indicating it became increasingly more difficult to detect IEDs. The experienced drivers fluctuated but had the easiest time identifying IEDs in segment two.

Segment	Average Sensitivity
2	1.39
3	1.01
4	0.90

Table 11: Average Sensitivity across Segments

Experience Level	Average Sensitivity
Novice	1.10
Experienced	1.18

Table 12: Average Sensitivity across Experience Level

	Average Sensitivity	
Segment	Novice Drivers	Experienced Drivers
2	1.39	1.39
3	1.06	0.98
4	0.48	1.16

Table 13: Average Sensitivity by Segment between Experience Levels

Signal Detection Theory Response Criterion

The response criterion, or β, for each driver was also calculated to analyze their IED detection performance. The response criterion, or bias, represents the likelihood of each participant to identify an item as either an IED or a non-IED. Equation 2 provides the equation used to calculate each participant's response criterion.

$$\beta = e^{-(Z(H)^2 - Z(1-FA)^2)/2}$$

Equation 2: Formula to Calculate Response Criterion

Table 14 shows the average response criterion for each participant across segments two through four. As previously stated, there is no response criterion calculation for segment one because there were no IEDs presented, so no hits were possible. It should be noted that Participant 12 only completed segment 1 and a portion of segment 2. The response criterion calculated for Participant 12 only reflects the small portion of segment 2 he completed.

Only 2 participants (1 and 12) had a β value greater than 1. This means that these participants engaged in more conservative behavior and they needed more evidence before they would call an item an IED. Participant 1 was classified as an experienced driver, with 8 years of service while participant 12 was classified as a novice driver with only 3.5 years of service.

Participant 12's response criterion only reflects his performance in a portion of one segment. He was not able to finish the course, so this response criterion is based solely on segment two. His true average may have been different had he been able to complete the entire course.

The other 12 participants all had average β values less than 1. These β values mean that the participants engaged in more risky behavior and required less evidence to identify an item as an IED.

Participant Number	Experience Level	Average Criterion
1	EXP	1.22
2	EXP	0.49
3	EXP	0.09
4	NOV	0.30
5	EXP	0.14
6	NOV	0.15
7	EXP	0.09
8	EXP	0.20
9	EXP	0.27
10	NOV	0.44
11	NOV	0.62
12	NOV	*3.55**
13	NOV	0.34
14	EXP	0.69

Table 14: Average Response Criterion for Each Participant (* *indicates an outlier criterion*)

Table 15 shows the average criteria across all participants for each segment. The criteria are all less than one, indicating that on average all participants had riskier behavior and did not require much evidence to call an item an IED. The criterion peaked in segment three, when overall the behavior tended to be slightly more conservative, but was the most liberal in segment four. Between experience levels shown in Table 16, both groups had risky criterion. The experienced drivers had a lower criterion value, indicating that overall, they were more liberal in their identification of IEDs than the novice drivers. When broken down by segment in Table 17, the experienced drivers were most conservative in their IED identification in segment three, while the novice drivers were most conservative in segment two. In general, both groups had criterion values less than one, indicating minimal evidence was required to identify an item as an IED.

Segment	Average Criterion
2	0.44
3	0.68
4	0.29

Table 15: Average Criterion across Segments

Experience Level	Average Criterion
Novice	0.90
Experienced	0.40

Table 16: Average Criterion across Experience Levels

	Average Criterion	
Segment	Novice Drivers	Experienced Drivers
2	0.83	0.14
3	0.33	0.90
4	0.49	0.16

Table 17: Average Criterion by Segment between Experience Levels

Eye Blink Frequency

The eye blink frequency for each participant was calculated from the eye tracker data. As participants navigated the course, blink frequency was expected to increase. In order for drivers to identify and comprehend more information, they would spend more time searching the screen, not fixating on particular locations for long periods of time. For several participants, significant portions of their eye tracker data were missing. This could be due to the participant moving their head outside of the calibrated zone for the eye tracker, or the equipment malfunctioning during data collection. Additionally, since each participant drove through the course at different rates, each participant's blink frequency was normalized to blinks per minute. Normalization of the data allowed the blink frequencies to be compared across all 14 participants. Blinks varied widely between participants, but within each participant there was little variability between segments. These average blink frequencies for each participant across the four segments are shown below in Table 18. Blank values for blink frequency in any particular segment meant that no blinks were captured during that segment. After an ANOVA was performed, it was concluded that eye blink frequency did not show a significant difference in workload between segments ($F(3,29) = 0.35$, $p = 0.786$). However, there was a significant difference between experience levels of the drivers ($F(1,29) = 5{,}49$, $p = 0.026$).

		Average Blink Frequency (blinks/min)			
Participant #	Experience Level	Segment 1	Segment 2	Segment 3	Segment 4
1	EXP	15.4		1.5	1.2
2	EXP	7.0	6.1	12.1	
3	EXP	11.3	9.6	11.7	
4	NOV	23.6	22.3	25.6	15.5
5	EXP	2.1	1.0		
6	NOV	0.3	0.9	0.5	0.6
7	EXP	28.3	39.3	36.6	45.8
8	EXP	0.5	0.5	1.1	2.5
9	EXP	6.7	7.5	9.6	8.6
10	NOV	2.6	4.7	4.3	5.8
11	NOV	5.0	6.4	11.2	11.2
12	NOV	7.6	2.0		
13	NOV	2.5	0.6	0.2	1.0
14	EXP	0.7	0.9	4.2	1.8

Table 18: Average Blink Frequency for Each Participant across Segments

Figure 14 shows the average blink frequency grouped by the two driver experience levels—novice and experienced. Although both groups had a very large range of blink frequencies, the two groups were significantly different. Experienced drivers had a higher average blink frequency across all four course segments than the novice drivers.

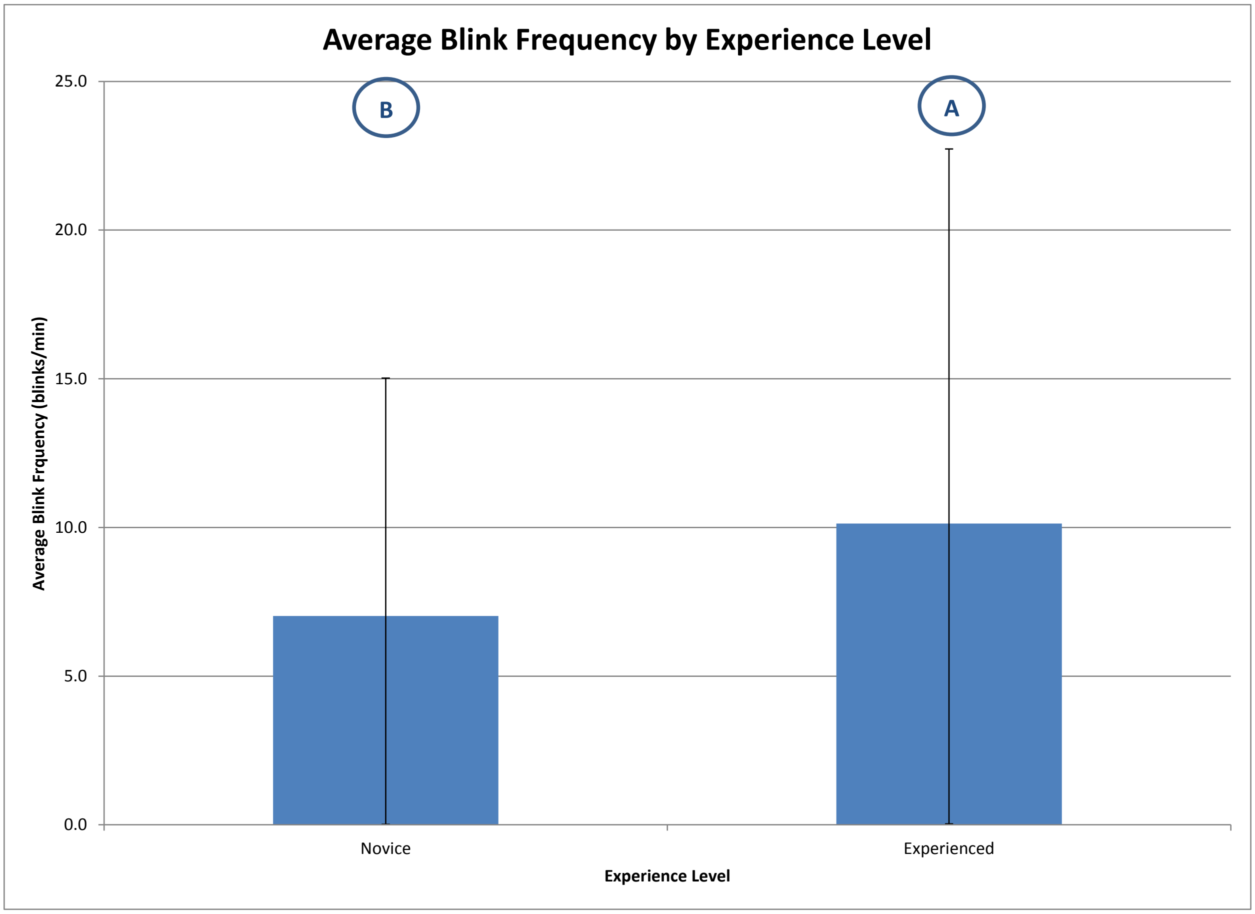


Figure 14: Average Blink Frequency by Experience Level

A two-sample t-test was performed on the blink frequency when it was broken down by segment for each individual participant. The blink frequencies were grouped by hits and false alarms in one category and by misses and correct rejections in another. Hits and false alarms both correspond to participants responding “yes” regardless of whether there is a target or a distractor. Misses and correct rejections mean that the participant responded “no” to either a target or a distractor. The intent was to identify if participants blinked more frequently when they responded “yes” to a situation versus responding “no,” regardless of whether their response was correct. Results from the two-sample t-test ($t(81) = 0.02$, $p = 0.988$) determined that there was no significant difference in the blink frequency when participants responded “yes” versus “no.”

Eye Blink Duration

Each participant's eye blink duration was also calculated from the eye tracker data. Shorter blink durations were expected as workload increased because participants would be scanning the screen more frequently looking for targets and absorbing their surroundings. Just like the blink frequency, there were some gaps in certain participants' data collection related to blink duration. Blink durations were very consistent between segments. The mean blink durations of segments 1, 2, 3, and 4 were 307.7 milliseconds, 323.3 milliseconds, 313.8 milliseconds, and 302.1 milliseconds, respectively. There was slightly more variation between experience levels, but not enough to be significant. Novice drivers had a mean blink duration of 287.4 milliseconds, and experienced drivers has a mean duration of 350.0 milliseconds. These average blink duration for each participant across the four segments are shown below in Table 19. After an ANOVA was performed, it was determined that eye blink duration also did not show a significant difference in workload between segments or experience levels of the drivers ($F(3,29) = 0.44$, $p = 0.728$; $F(1,29) = 0.18$, $p = 0.671$, respectively).

		Average Blink Duration (ms)			
Participant #	Experience Level	Segment 1	Segment 2	Segment 3	Segment 4
1	EXP	256.6		278.1	243.8
2	EXP	364.0	349.5	341.1	
3	EXP	437.7	378.2	342.2	
4	NOV	319.7	344.6	314.9	302.7
5	EXP	378.5	375.4		
6	NOV	216.9	318.0	367.0	355.4
7	EXP	261.1	271.5	274.2	269.6
8	EXP	242.8	314.9	323.1	280.3
9	EXP	287.7	289.9	278.2	301.7
10	NOV	336.6	294.0	317.0	315.7
11	NOV	354.0	318.5	371.8	337.5
12	NOV	295.2	300.3		
13	NOV	276.9	302.2	258.6	279.1
14	EXP	279.5	345.6	309.5	335.1

Table 19: Average Blink Duration for Each Participant across Segments

A two-sample t-test was performed on the blink duration data when it was broken down by segment by individual participant. The blink durations were also grouped by hits and false alarms and by misses and correct rejections. The expectation was that blink duration would be lower when participants responded “yes” to a situation versus responding “no.” It was determined that there was no significant difference ($t(81) = -0.05$, $p = 0.961$) in the blink duration when participants responded “yes” versus “no.”

EEG Channel Time-Frequency Analysis

Each participant’s brain activity was collected with electroencephalography (EEG). Changes in frequency were displayed on a channel time-frequency graph with event-related spectral perturbation (ERSP). Darker, more intense colors, such as red, orange, and yellow indicate increased brain activity and were expected to increase during the course. Time-frequency analyses were run for all 14 participants on 6 EEG channels. These six frontal midline EEG channels were selected because this area of the brain is known to display changes in workload levels. For each EEG channel, two frequency levels were analyzed—alpha (8-13 Hertz) and beta (13-30 Hertz). Changes in intensity were then compared to the event log to identify which events triggered changes in EEG activity. Samples of the time-frequency analyses are shown for selected participants in the figures below to highlight the changes in activity observed. An analysis of the intensity of the brain activity was performed on all 14 participants. The results below are presented with a summary analysis first and then individually by participant. The times identified for each participant below indicate the time frames where increased brain activity was present in the time-frequency graphs. All time-frequency analysis images can be found in Appendix F to observe the trends noted at the time periods described for each participant.

Overall EEG Results

Appendix G is provided as a summary result to the individual EEG results presented in the sections above. The first column labels the four segments that were encountered during the experiment. Therefore, moving down each row corresponds to a later time in the driving course. The next two columns indicate any communications activity, targets or distracters that were encountered. It is important to note that this figure does not reflect the entire driving course, only periods of activity that should have

elicited changes in brain activity. The remaining columns for the 14 participants show times where the time-frequency analysis indicated changes in brain activity (highlighted in yellow) and their responses to targets and distracters. The yellow coloring here only indicates significant changes in EEG activity identified in the previous sections; it does not reflect the intensity of their EEG activity.

Elevated brain activity was expected for essentially the entirety of the figure, especially in segment four. The communications activity were times when the participants had to either acknowledge to a target or distracter or answer questions that required cognitive processing and decision making. EEG should have been sensitive to these changes in cognitive processing especially if the participants were experiencing elevated workload levels.

It is interesting to note how some subjects exhibited changes in activity for some stretches and then fell “silent” during other times when higher workload was expected. Participant one especially has a large amount of activity in segments two and three but then exhibits no changes in activity in segment four. These trends are inconsistent with the subjective MCH ratings that exhibited an overall increasing trend in workload progressing from segment one to segment four. These EEG results exhibit more sporadic periods of activity and not an increasing trend of workload from segment one to four.

Participant One

Several time periods indicated an increase in brain activity corresponding to increased workload for participant one. 750,000 to 1,000,000 milliseconds correspond to the first half of segment 2, and there was increased brain activity as 2 IEDs were missed during that time. Between 1,250,000 and 1,750,000 milliseconds—the remainder of segment 2—this participant failed to identify 3 additional IEDs. One IED was correctly identified. Segment 3 lasted from 2,000,000 to 2,250,000 milliseconds. During that time, seven IEDs were missed and two distracters were incorrectly identified as targets. Channels 38 and 46 had some of the stronger readings and displayed these trends.

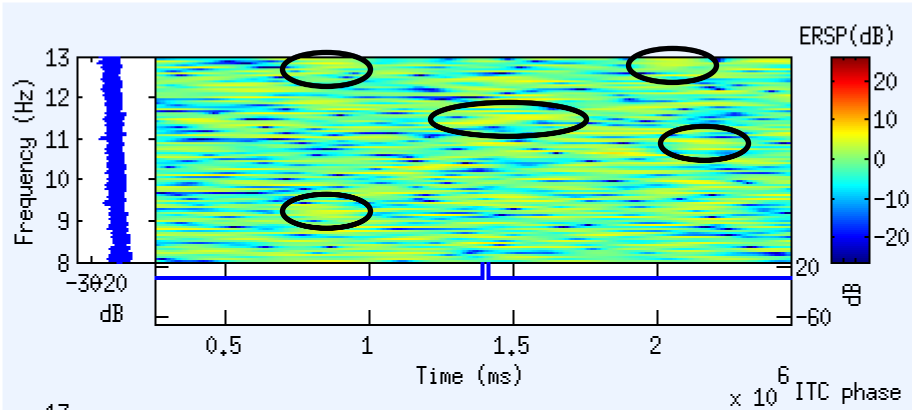


Figure 15: Participant 1 EEG Channel 46 Alpha Frequency

Participant Two

Channels 39 and 46 exhibited the strongest changes in brain activity for participant 2. In segment 1 up to 500,000 milliseconds, there was one distracter presented. The participant incorrectly identified it as an IED. The second half of segment 1 lasted from 600,000 to 850,000 milliseconds. During that time, there were two additional distracters presented and the participant correctly identified both of them as distracters. From 1,400,000 to 1,600,000 milliseconds, in the middle of segment 2, 3 IEDs were presented. The participant failed to identify two of these IEDs. During that same time frame, three additional distracters were presented and one was incorrectly identified as an IED. Lastly, during the end of segment 2 and beginning of segment 3, from 1,900,000 to 2,200,000 milliseconds, several distracters were presented and only 1 IED. The IED was correctly identified and all the distracters were either identified correctly or ignored.

Participant Three

There were three timeframes of interest for participant three. During the end of segment 1 and into segment 2 (800,000 to 1,200,000 milliseconds), only 1 IED was presented. This target was correctly identified by the participant. Additionally, four distracters were encountered. Two of these distracters were acknowledged and the other two were ignored. Between 2,200,000 and 2,400,000 milliseconds, 3 IEDs were encountered and only 1 was missed. All distracters presented during this time frame were all correctly identified. Lastly, from 3,250,000 and 3,600,000 milliseconds, 2 IEDs were presented. The first

one was missed but the second one was correctly identified. This time frame captured the end of segment four, just before the final ambush. The two channels that reflected the strongest changes in brain activity for participant 3 were channels 38 and 46.

Participant Four

Channel 38 for participant 4 showed some of the strongest changes in EEG activity. From 0 to 500,000 milliseconds, participant 4 encountered a distractor that was identified as an IED. After this event, a wrong turn was made on the course and the participant had to figure out where to go to get back on course. Between 1,400,000 and 1,600,000 milliseconds, the participant encountered 3 IEDs and 3 distractors. Two out of the three distracters were correctly identified correctly, but 2 of the 3 IEDs were missed. Finally, from 1,750,000 to 2,000,000 milliseconds, another change in brain activity was observed. This was the end of segment two, and concluded with a small ambush. The participant's performance was very accurate during this portion of the segment, correctly identifying all three IEDs and all four distracters presented. Participant 4 is missing the remaining 19 minutes of EEG data beyond segment 2.

Participant Five

Participant five had several periods where increased brain activity likely indicated increased workload. The two most prominent channels to observe this activity were channels 4 and 46. Shortly into segment 1 (250,000 to 600,000 milliseconds), 1 distracter was encountered. The participant correctly identified this distracter. Moving into segment 2, from 1,000,000 to 1,250,000 milliseconds, 1 IED and 1 distracter were encountered. Again, both items were correctly identified. There were 3 targets and 3 distracters encountered between 1,400,000 and 1,600,000 milliseconds. Only one IED was correctly identified. Additionally, one of the distracters was mistakenly identified as an IED. During the beginning of segment 3 (2,000,000 to 2,250,000 milliseconds), the sole IED encountered was correctly identified. Out of the other three distracters presented, one was correctly rejected and the other two were ignored. The last period of interested lasted from 2,500,000 to 2,700,000 milliseconds. During this time, three IEDs were presented and only one was missed. Two distracters were also presented. Unfortunately, the EEG data during the 19 minutes of segment 4 were not captured.

Participant Six

Channels 38 and 47 captured some of the most intense changes in EEG activity for participant 6. He exhibited many short but more frequent changes in workload. The first peak was noticed at the very beginning of the course up until 500,000 milliseconds into segment 1. Here only one distracter was encountered and it was correctly identified. Between 750,000 and 1,250,000 milliseconds, 5 more distracters were encountered. Four out of the five distracters were correctly identified and the other one was not acknowledged. There was only one IED encountered and it was correctly identified. The level of activity began to increase and between 1,750,000 and 2,200,000 milliseconds, participant 6 encountered 5 IEDs and 4 distracters. The participant failed to identify only one of these IEDs. The participant entered into the final segment around 2,600,000 milliseconds. Between 2,600,000 and 2,800,000 milliseconds, 4 IEDs were presented. The participant failed to identify three of these. Even though there were not many targets or distracters presented at this time, there was a lot of communication required on the part of the participant to the rest of the vehicles in the convoy, likely causing a distraction. Moments later (2,900,000 to 3,100,000 milliseconds), 2 more IEDs were encountered and also not noticed by the participant. The last change in brain activity occurs between 3,200,000 and 3,400,000 milliseconds. During this time, the participant identifies the last IED encountered just before the final ambush attack.

Participant Seven

There were four time periods with increased brain activity for participant seven. Up to 500,000 milliseconds in segment 1, the participant encountered and correctly identified 1 distracter. Between 1,000,000 and 1,250,000 milliseconds there was increased brain activity as 1 IED was encountered but labeled a non-threat. An additional distracter was also encountered and correctly identified. Between 1,300,000 and 1,700,000 milliseconds, the participant encountered 6 IEDs and 3 distracters. This participant failed to identify two of the IEDs. The other four IED was correctly identified. One of the distracters was also labeled an IED. In segment 3, from 2,100,000 to 2,500,000 milliseconds, two IEDs were encountered and both were missed. The one distracter was correctly identified. Channels 11 and 39

had some of the stronger readings of these trends. The EEG data in the entirety of segment 4 (10 minutes) were not captured.

Participant Eight

Channels 4 and 46 had the most prominent changes in brain activity for participant 8. In segment 1, between 500,000 and 750,000 milliseconds, there was one distracter presented and it correctly identified as such. From 1,100,000 to 1,300,000 milliseconds there was 1 additional distracters presented and the participant correctly identified it. From 1,500,000 to 1,750,000 milliseconds, 4 IEDs were presented. The participant only missed identifying one of these IEDs. During that same time frame, four additional distracters were presented and again they were all identified correctly. Lastly, for the entirety of segment 3, between 2,250,000 and 2,600,000 milliseconds, there were many targets and distracters encountered. Out of the seven IEDs encountered, five were correctly identified. The other two failed to be noticed. There were also 10 distracters presented. Only one was mistakenly identified as an IED. The 21 minutes of segment 4 did not have recorded EEG data.

Participant Nine

Participant nine's more intense brain activity does not begin until the second segment. From 1,100,000 to 1,400,000 milliseconds, 3 distracters were presented and all were correctly identified. Then, between 1,400,000 and 1,600,000 milliseconds, 3 IEDs were encountered and 2 were missed. Of the two distracters presented, both were correctly identified. Next, from 1,600,000 and 1,800,000 milliseconds, 3 more IEDs were presented along with 3 distracters. All six items were acknowledged and correctly identified. Moving into segment 3, there were 3 IEDs and 3 distracters encountered between 2,000,000 and 2,250,000 milliseconds. The IED was again correctly identified. Regarding the distracters, two were correctly identified as distracters and the other two were labeled as IEDs. Lastly, in segment 4, there were 11 items encountered between 2,500,000 and 2,800,000 milliseconds. Eight of these items were IEDs and the participant failed to identify three of them. The two distracters were all acknowledged and correctly identified. The two channels that reflected the strongest changes in brain activity for participant 9 were channels 11 and 38.

Participant Ten

Channel 46 showed some of the strongest changes in EEG activity for participant 10. 750,000 and 1,000,000 milliseconds was the beginning of segment 2. The participant correctly identified the two distracters presented. Between 1,200,000 and 1,400,000 milliseconds, the participant encountered 5 IEDs and 5 distractors. Three of the IEDs were not identified and one of the distractors was misidentified as an IED. There was a lot of information presented to the participant in this small window of time. 1,500,000 to 1,600,000 milliseconds was the beginning of segment 3. The participant was presented with three distractors but identified two of them as IEDs. Finally, between 200,000 and 2,250,000 milliseconds, the participant encountered 6 IEDs and 3 distractors in this portion of segment 4. Only two of the IEDs were missed.

Participant Eleven

Participant 11 also had several time periods where changes in workload were observed. The most prominent channels to feature these changes were channels 38 and 39. During the end of segment 1 and into the beginning of segment 2 (600,000 to 800,000 milliseconds), there was 1 IED encountered and it was correctly identified. From 1,300,000 to 1,600,000 milliseconds, 6 IEDs and 5 distracters were presented. The participant missed identifying 2 of the IEDs. Additionally, one of the distracters was mistaken for an IED. Participant 11 drove through the majority of segment 3 between 1,750,000 and 2,000,000 milliseconds. There was a lot of information presented during this time and the participant performed rather well. Out of the six distracters encountered, two were mistaken for IEDs. Only one actual IED missed being identified; the other three IEDs were identified correctly. The only IED presented during 2,200,000 and 2,400,000 milliseconds was missed. Additionally, two of the four distracters encountered were incorrectly identified as IEDs. Lastly, from 2,600,000 to 3,000,000 milliseconds, 2 IEDs were encountered and correctly identified. The final ambush began at the end of the peak in brain activity—around 3,000,000 milliseconds.

Participant Twelve

Participant 12 only had EEG activity through the beginning of segment 2 when the experiment was ended early for this individual. This was the only participant to exhibit shades of orange on the ERSP scale, indicating that this participant experienced greater levels of workload than the other participants. This feature is most apparent on channel 38, shown below in Figure 16. There were other areas of interest besides the end of the segment. Up to 200,000 milliseconds, there were no targets or distracters presented; however, the driver made a wrong turn and had to turn around to get back on course. Between 350,000 and 500,000 milliseconds, the distracter presented was incorrectly identified as an IED. Still in segment 1, between 800,000 and 1,000,000 milliseconds, there were 3 additional distracters encountered. One was mistakenly identified as an IED and the other two were correctly rejected. The orange area highlighted in the figure below occurs between 1,200,000 and 1,400,000 milliseconds. Here, participant 12 entered into segment 2. One IED was encountered and correctly identified. Out of the three distracters presented, the last one was incorrectly identified as an IED and then the participant chose to end the experiment.

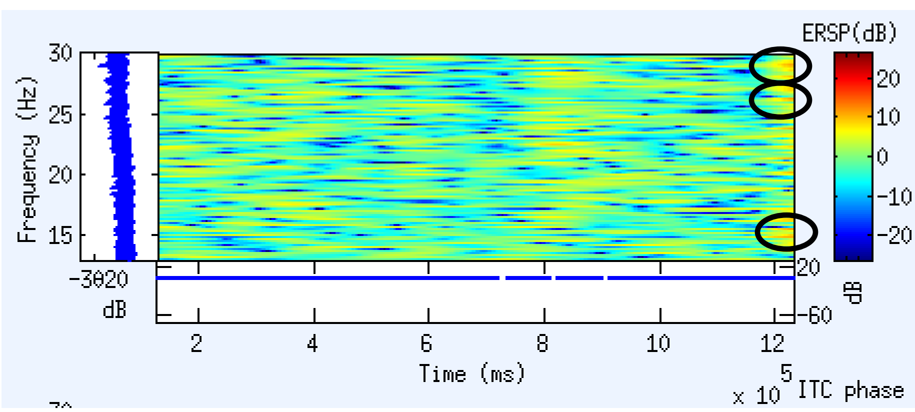


Figure 16: Participant 12 EEG Channel 38 Beta Frequency

Participant Thirteen

Channels 38 and 46 featured some of the more noticeable changes in EEG activity for participant 13. During the latter half of segment 1 (600,000 to 1,000,000 milliseconds), this participant was presented with 3 distracters. Two were identified as such and the third was ignored. Only 1 IED was encountered between 1,250,000 and 1,500,000 and it was correctly identified as such. The other three distracters presented were all acknowledged; however, one was mistakenly identified as an IED. About 2/3rds of

segment 3 was captured between 2,100,000 and 2,500,000 milliseconds. During this time, many targets and distracters were experienced. 10 distracters were presented and all were correctly acknowledged excepted for 1 which was called an IED. Only two IEDs were presented—one was correctly identified and one was missed. Lastly, from 2,600,000 and 3,000,000 milliseconds, 2 out of the 5 IEDs encountered were missed. Six distracters presented and only one was incorrectly identified as an IED. One was ignored and the other four were correctly acknowledged.

Participant Fourteen

For participant 14, channels 46 and 47 exhibited the most noticeable changes in EEG activity. The first area of interest lasted up to 500,000 millisecond during segment 1. Here, two distracters were encountered and both were correctly acknowledged as such. Between 750,000 and 1,000,000 milliseconds, 4 more distracters were encountered. All were acknowledged, but one was mistakenly identified as an IED. The sole IED presented was not identified. The latter half of segment 3 was captured between 1,400,000 and 1,800,000 milliseconds. During this time, three out of the four IEDs encountered were identified correctly. The three distracters presented were either acknowledged or ignored. The busiest section that elicited changes in brain activity for participant 14 lasted between 2,000,000 and 2,250,000 milliseconds. The number of targets and distracters was not any more than previously encountered; however, there was more communications activity. Again, three out of the four IEDs were correctly identified. This time, two out of the three distracters were mistakenly identified as IEDs.

Prediction of Miss Rates

The scatterplots for each of the explanatory variables was created to visually assess the impact on the participants' miss rates. The scatterplot with the strongest relationship is shown below in Figure 17. The other scatterplots can be found in Appendix E.

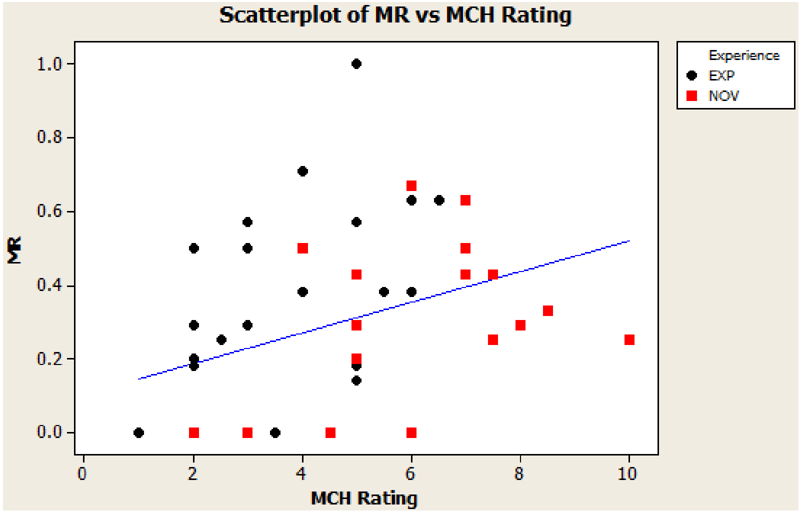


Figure 17: Scatterplot of Miss Rate versus MCH Rating

As the course progressed, it was expected that the miss rate would increase as participants experienced additional workload. As predictors, blink frequency and MCH rating should increase with higher miss rates while the blink duration should decrease. The initial linear regression results showed that 25.5% ($F(3, 36) = 4.10$, $p = 0.013$) of the variability in the miss rate could be explained by these three predictors. There was inconclusive evidence to determine that blink duration was a significant predictor and not zero ($t(36) = -0.56$, $p = 0.579$).

The new linear regression with only MCH rating and blink frequency as predictors of miss rate explained 20.6% ($F(2, 42) = 5.46$, $p = 0.008$) of the variability. Both MCH ratings and blink frequency proved to be significant predictors of miss rates ($t(42) = 2.82$, $p = 0.007$ and $t(42) = 2.12$, $p = 0.040$, respectively). As initially expected, both coefficients were positive. This indicates that as blink frequency and MCH ratings increase, the miss rate increases, indicative of elevated workload.

Discussion

The hypotheses developed for this thesis were centered on three main factors: the experience level of the drivers (novice versus experienced drivers), the workload assessment of different metrics (physiological versus subjective), and task performance. Regarding the experience level of the drivers, it was hypothesized that participants who were novice drivers would encounter higher levels of perceived workload than expert drivers. For the workload assessment using different metrics, it was hypothesized that blink frequency and electroencephalography measures of mental workload would be correlated. Additionally, it was hypothesized that significant differences in average blink frequency between the four course segments would correspond to differences between subjective workload ratings for the four segments. For task performance, it was hypothesized that as mental workload increased, the miss rate and false alarm rate for IED detection would increase.

It is important to note that certain technical issues prevented following the original 2 x 2 within-subjects design of the experiment. Recall, data collection efforts were not part of this thesis. As previously mentioned, not all participants were exposed to the same communications activity throughout the experiment. The resulting workload from the amount of communications likely varied among participants. Consequently, this had some impact on the statistical significance of the experiment. This reduced significance compounded with the already small sample size compromised some statistical analysis of the data. Statistical conclusions were able to be gained from the signal detection and subjective rating data. In future experiments, more care should be taken to ensure the quality of the data to draw statistically significant conclusions.

Driver Experience

The experimental results supported the first hypothesis, which theorized that novice drivers would have a higher perceived workload than experienced drivers. When an analysis of variance was performed on the data, the Modified Cooper Harper ratings were significantly lower for experienced drivers (Figure 9).

It is important to understand that less experienced drivers are going to perceive a higher amount of workload than more experienced drivers. Higher perceived workload can lead to elevated stress levels and performance decrements. Recognizing these workload differences related to experience could lead to changes in training and scheduling during deployments. Some actions can be taken to help mitigate or minimize overloading especially to less experienced MRAP drivers. While drivers are still gaining experience, they could be tasked to drive in areas that are not necessarily known for many IED attacks. It should be recognized that less experienced drivers are going to feel more pressure in difficult situations and may have performance decrements due to the increased perceived workload they experience.

Workload Assessment Metrics

Eye Blink Frequency

The experimental results did not support the hypothesis that there would be significant differences in blink frequency between the four course segments. Literature has shown that events create arousal, resulting in increased blinking (Fukuda, Stern, Brown, and Russo, 2005). However, with respect to each individual participant, the average blink frequency was nearly the same across all four course segments (Table 18). The analysis of variance confirmed the observation that the blink frequencies were not significantly different between the four segments (Figure 14). However, the blink frequency was significantly different between the experience levels of the participants. This result would suggest that the experienced drivers were subjected to a higher workload level than the novice drivers.

There are several reasons why the blink frequency results indicated that experienced drivers had a higher workload level than novice drivers. It should be noted that the experienced drivers had more misses but fewer false alarms than the novice drivers. It is possible that the experienced drivers initially missed identifying an IED, but once they drove up closer to it and realized the item was an actual IED, the consequences of their miss in reality set in. This could have led to an increased level of stress and perhaps an elevated workload level.

Another explanation for the increased blink frequency of experienced drivers is their awareness of their surroundings. Experienced drivers are likely more aware of their surroundings, causing them to

blink more. They could spend more time scanning and searching the screen for possible threats and other items of interest along the course. Novice drivers could have been fixating more on particular areas of the screen, not fully capturing all the information provided to them.

There were additional performance metrics captured for the overall study but outside the scope of this research. The novice drivers could have allowed other aspects of their performance to lapse, but it was not captured in this study. One performance metric that may have been compromised due to increased workload is driving performance. Certain participants could have allowed their driving performance to suffer in order to maintain a low workload level. However, driving performance is outside the scope of this research.

Since many participants' eye blink data were missing significant portions of data, it is important to interpret these results cautiously. Additionally, it is important to follow up this study with additional research. Using an upgraded eye tracker system could eliminate some the problems experienced in this data collection effort. A follow-up study is recommended so that better eye blink data can more accurately determine the true workload experienced by MRAP drivers.

Eye Blink Duration

Another eye metric analyzed was eye blink duration. It was expected that blink duration would decrease with increasing workload. As shown in Table 19 and confirmed in the ANOVA, there were no significant differences between the four course segments or the two experience levels of the participants. Although no results proved to be significant, it is interesting to note how experienced drivers had longer blink durations than the novice drivers. This finding is somewhat contradictory to the findings from the blink frequency analysis because the experienced drivers were found to have higher workload levels, indicating their blink durations should have been shorter. Again, a follow-up data collection effort is recommended to more accurately assess the workload measured through an eye tracker system.

Eye Metrics by Event

Since the blink frequency and blink duration metrics proved to be mostly inconclusive at a low level, the eye blink data were analyzed at a higher level of detail. Two-sample t-tests were run to

determine if there was simply a difference between participants responding “yes” to an item on the screen versus responding “no.” While the eye blink data were generally insignificant at the event level, these tests tried to differentiate if the eye would even change when a driver suspected a target versus not. Again, because of the large amount of missing data, it was hoped that at a very high level, some trend would be apparent. Unfortunately, the t-tests for both blink frequency and blink duration proved to be insignificant, concluding that there were no differences in the eye metrics when an individual simply responded “yes” or “no.”

EEG

Across all participants, the six different EEG channels showed similar trends in regards to changes in brain activity during the experiment. As noted before, there were some channels that showed more noticeable changes in this activity. In general, channels 38 and 46, found in Figure 4, often showed the most prominent changes in brain activity; however, every channel was prominent for at least one participant.

While most of the changes noted were in the yellow region of the ERSP, participant 12 did have some orange coloring just before the experiment was ended early. The more intense orange coloring indicates that this participant was experiencing rather high workload combined with simulator sickness that resulted in the experiment being cut short. There were three participants—9, 13, and 14—who had less intense changes in EEG activity according to the ERSP. The intensity of their ERSP coloring was not as dark as the other 11 participants. It is possible that these participants did not perceive the situation to be as challenging or intensive as did the other participants.

As noted from the previous EEG results and the summary in Appendix G, it was often the case that EEG activity would change when the participant missed an IED or made a wrong turn on the course. It is not surprising that these situations would be more stressful and cause the participant to experience higher workload. The figure provided in Appendix G also illustrates the selective nature of EEG in this experiment. While EEG was able to capture certain changes in workload from particular events, there were also many times where the participant’s tasking did not exhibit any change in workload according to

the EEG recordings. It is unclear exactly why EEG appears to be selective as to which events actually attributed to noticeable changes in workload.

These results show the limitation or weakness of EEG data. It is highly sensitive to outside factors which can make it an unreliable metric for workload assessment. Even with experienced data collectors, EEG can still provide inconclusive results, requiring large amounts of time to collect, process, and interpret the data. Utilizing another analysis technique on this EEG data or a more experienced EEG rater may provide different results. As with the eye tracker data, additional research is required before recommendations can be made regarding the use of EEG to assess mental workload.

Eye Blink Frequency and EEG

Due to the fact that the EEG data interpreted from this experiment was qualitative, a correlation between these results and the results from the eye blink frequency was not able to be performed. At this time, the approaches to quantitatively analyze EEG data are still very theoretical and not applicable for this study. Additionally, the eye blink frequency did not produce any statistically significant results. The EEG data were able to identify trends in each participant’s perceived workload during the experiment.

Modified Cooper Harper Rating

Figure 3 indicates that the course segments became progressively more challenging. As rated subjectively by the participants, overall the course progressed in difficulty from segment one to segment four. Although the participants subjectively felt that the course increased in difficulty, their eye metrics were not consistent with this finding. The eye metrics across all four segments indicated that there was no change in difficulty across the course segments. The results found with the simple MCH scale make sense and are not subject to the sensitivities of the other methods. MCH also supported the hypotheses that were posed about the effects of experience and task difficulty on workload.

The difficulties with the eye tracking data could have contributed to the insignificant eye tracker results. It is hard to draw accurate conclusions about the compatibility between the eye tracker results and the subjective results to understand if the participants were able to accurately assess their workload during the driving course. It is also important to note that in a simulator, there are far fewer distractions than

faced in a true operational environment. These distractions could come from other crew members, the GFE placed in the system, and other real-world elements. The conflicting results as to the true level of workload experienced could be resulting from the far less distracting environment of the simulator as compared to the battlefield. The simulator may not have provided a comparable level of workload and pressure that is regularly experienced while on deployment.

Prediction of Miss Rates

A linear regression analysis was performed in an attempt to explain participants' miss rates through three predictor variables—MCH rating, blink frequency and blink duration. It was expected that the MCH rating and blink frequency would increase with elevated miss rates, while the blink duration would decrease. The linear regression indicated that only MCH ratings and blink frequency were significant predictors of increased miss rates. Additionally, these two variables only explained approximately 20% of that variability. Since the blink data were lacking overall quality, a linear regression with solely MCH ratings as a predictor was run. Figure 37 in Appendix E shows that 14.1% ($F(1, 48) = 7.89$, $p = 0.007$) of the variability can be explained by this one variable alone. Although none of these predictors explain a large portion of the miss rates, it is possible that with more complete data, additional explanatory power might be gained. At this time, it is not recommended to use MCH ratings and blink frequency to attempt to determine the number of IEDs a participant will miss. However, it does highlight previous findings that increased workload will negatively affect driver performance in regards to detection tasks.

Task Performance

The most effective tool for detecting IEDs is a trained soldier's naked eye. The extensive training materials developed by the military are the best way to prepare soldiers for the wide variety of environments they may face to detect IEDs once deployed. The trainings include the use of scenarios, simulated environments, and field exercises. Videogame-based training already exists as one tool to help prepare soldiers for deployment and train them to accurately detect IEDs. The course that soldiers experienced in the simulator for this research is likely similar to previous training they received.

IED Miss Rate and False Alarm Rate

The fourth hypothesis was not confirmed with the findings of this study. It was hypothesized that as mental workload increased, the miss rate and false alarm rate for IED detection would increase. However, there were no significant differences in the miss rates or false alarm rates for the participants across the four segments.

The course was designed to increase in difficulty. The subjective ratings indicated that the course increased in difficulty. The limited amount of blink frequency data suggested that the course did not subject the workers to increasing difficulty. It is unclear exactly how much more difficult the course became due to the differences in the results of these metrics. If the course became more challenging, but was not overwhelming or comparable to amount of workload they face in the real environment, then it is not surprising that their performance did not decline. It appears that the Modified Cooper Harper scale is more sensitive than the blink data; however it is hard to accurately conclude this given the quality of the blink data.

Since the course was designed to increase in difficulty, it is possible that the results suffered from the impact of the order effect. The step functions found for the MCH ratings progressing from segments one through four could have been a result of this effect. If the participants recognized or knew that the course was becoming more difficult, their MCH ratings could have been impacted as they knew to rate the successive segments slightly more difficult. To combat the order effect, the presentation of the segments should have been randomized. This would have eliminated the effect because the participants would not know if a successive segment was easier or more difficult than the previous.

Warm, Parasuraman, and Matthews (2008) found that detection performance declines over time. These decrements are usually seen within the first 15 minutes of a detection task. Novice individuals are not the only ones who experience attention decrements; they can occur in experienced individuals too. Although it may be suspected that detection tasks are often understimulating because individuals are searching for infrequent signals, these tasks are in fact very resource demanding and cause the individual to experience high workload. Performance declines with time on task. This task only lasted on average 55

minutes (depending on how quickly they drove though the course). It was likely not long enough to really have a lapse in attention or onset of fatigue.

The military has also studied trends related to casualties from failed IED detections. A report by Leipold (2009) found that the majority of IED casualties occurred at the beginning and end of deployment. The casualties at the beginning often resulted from a lack of familiarity with the surroundings. The casualties at the end were a result of soldiers focusing on their return home. The report also found that some soldiers had an innate ability to detect IEDs. Extensive training was also shown to be a valuable tool to help train one's detection ability to keep Warfighters safe during their deployments. The most important trait is to remain focused so that lapses in attention or vigilance are minimized.

Although the false alarm rates were not different between the course segments, it is interesting to note that the false alarm rate was significantly lower for experienced drivers than novice drivers (Figure 13). This result indicates that the novice drivers were more liberal in identifying potential IEDs. The novice drivers were more likely to identify an item as an IED when they may not have had substantial proof that it actually was one.

It is possible that due to the small data set (N = 14), these IED results are simply a result of statistical phenomena from the small sample size. However, these results could be related to a difference in training. During training, MRAP drivers are instructed to identify every potential threat as an IED. The novice drivers are closer to following protocol in identifying any potential threat as an IED, hence the higher false alarm rate. The experienced drivers potentially interjected more personal experience and subjectivity into the IED detection. These differences would indicate that training is highly relevant to how drivers attempted the course. The statistical differences could be a result of the differences in the way the drivers attempted the course. It is possible that the novice drivers were more focused on their performance and scoring on areas where they knew they would be measured. However, the experienced drivers may have been more concerned about their overall performance across the entire study. Future studies are recommended in order to identify the specific causes and changes to driving performance.

Signal Detection Theory Sensitivity

There were some interesting trends observed with regards to the sensitivity and response criterion from the signal detection analysis. It was found that the average sensitivity decreased between segments. Initially, the second segment had the highest sensitivity indicating that it was easiest to distinguish the IEDs from all other distracters. By the fourth segment, the sensitivity had decreased, and it was not as easy for the participants to distinguish the IEDs.

As the segments progressed, there were increased vehicle communications and more distractions. Sensitivity indicates how easy or difficult it is to distinguish a target from distracters. One would expect that as workload and distracters increased, it would be more difficult to distinguish the targets with a high degree of accuracy. It is logical that the sensitivity would decrease and it would be more difficult for participants to distinguish the targets from the distracters.

Although the sensitivity decreased, it did not affect the drivers' performance as far as accurately detecting IEDs. As previously mentioned, the miss rates were not significantly different between the latter three segments. Although the IEDs were more difficult to distinguish, participants were still able to accurately identify and report the IEDs.

While driving performance remained steady, driver workload increased across the segments. It is possible that the decrease in sensitivity was one contributing factor to the increase in mental workload. As the sensitivity decreased and it became harder for the drivers to distinguish targets from the surrounding distracters, their workload increased as they continued to keep their detection performance at a high level.

Signal Detection Theory Response Criterion

It was found that the average response criterion peaked at segment three and was the lowest at segment four. Across all participants, the average β value for each segment is less than one. Overall, most of the participants are engaging in "risky" behavior when identifying items as IEDs; that is, they require little evidence to identify an item as an IED. The third segment had the highest response criteria indicating that slightly more evidence is required for identification. The average criterion is the lowest in segment four, likely because of all the other distractions that are present in that segment. Participants do

not want to take the chance of missing an actual IED, so they do not require much evidence to classify an item presented as an IED.

Between experience levels, the experienced drivers had a lower response criterion, although both experience levels were considered to have “risky” criterion. These results would disagree with the miss rate and false alarm rate results, since the experienced drivers had a higher average miss rate and lower average false alarm rate. Table 14 shows that Participant 12’s response criterion was a large outlier for the novice drivers. Removing that criterion from the novices’ average, the response criterion for the novices falls to 0.37. Without the outlier, the results confirm the miss rates and false alarm rates.

There are factors that are known to influence the response criterion. Knowledge regarding the proportion of trials which have a signal and real or perceived costs are two factors that affect response criterion (Wickens, 2002). When there are costs or values associated with alternatives, the observer will attempt to adjust their criterion to minimize cost or maximize gain. In the situation of IED detection, where misses are costly and false alarms are less of a concern, the driver should shift to reduce the miss rate. While there were no actual bomb attacks or explosions in the simulator, many of the participants treated their drive through the simulator as though it was an actual mission. In some instances, the conditions in the simulator were enough to trigger post-traumatic stress disorder (PTSD). Even in a laboratory setting without real attacks or IED detonations, the participants had a tendency to be more conservative in their classification of potential threats, meaning less evidence was required to classify a target. In this research, it is difficult to assess if the cost of misses was enough to change the participants’ response criterion, since baseline criteria were not established prior to the experiment.

Relevance of Experiment

The results from this laboratory experiment have real-world implications. Regardless of the type of metric chosen—subjective or physiological—one metric cannot provide a complete understanding of the workload experienced. In this experiment, the subjective Modified Cooper Harper ratings were as good, if not better than the objective biosignals. The MCH ratings produced consistent and expected results. The eye blink data produced insignificant and conflicting and inconclusive results, potentially

because of the serious issues with data collection. The EEG data also produced conflicting results, noting at times how the data appeared to be very selective as to which periods of experiment activity actually resulted in noticeable changes in brain activity. Additionally, both the eye blink and EEG data produced results that were not expected. Given the reasonable performance of the subjective rating data and the overhead cost of EEG and eye blink data collection—time, expense, and expertise to administer, collect, and analyze—the expense of physiological tools may not be justified.

This study utilized a highly controlled environment to attempt to analyze the differences in workload through the use of EEG and eye blink data. Even in this environment, there were difficulties collecting robust data. Before extending this type of experiment into a more operational setting, these data collection challenges would need to be addressed and overcome in order to collect useful data for analysis. The one piece of promising data comes in the form of the subjective MCH ratings. In a simulation environment, this tool was able to provide useful results and perhaps could be used in a more authentic environment. The subjective ratings did not appear to interfere with overall task performance. There will always be limitations in a field environment, but where there is risk of instrumentation interfering with subject, MCH does not interfere.

Although this thesis was not responsible for these data collection efforts, some recommendations can be made regarding additional research. In future studies, care should be taken in order to ensure that each assessment metric is collected properly and throughout the entirety of the experiment. For this research, better eye tracking software and data would likely have produced different results. Additionally, EEG analysis is a very specialized field and there are many different analysis techniques that can be used. For this experiment, a simple technique was used. Future studies may choose to have more experienced EEG raters or utilize another technique to interpret the results.

Conclusions

This research analyzed the effect that mental workload has on MRAP vehicle driver performance. The differences in different workload assessment techniques were also studied as part of this research. Participant workload was measured subjectively with the Modified Cooper Harper rating scale and physiologically with electroencephalography, blink frequency, and blink duration. The different metrics were selected to compare their sensitivity and ease of use in an operational environment. Task performance through IED detection was also studied as potential differentiating factors between the two driver experience levels. Fourteen Marine Corps heavy vehicle drivers were studied as they participated in a simulated convoy mission in the Ride Motion Simulator.

It was originally hypothesized that blink frequency and subjective workload ratings would complement each other and each show significant differences in workload between the four course segments. It was also theorized that novice drivers would experience higher levels of workload than experienced drivers. The findings of this research partially support the original hypothesis. There were significant differences in the subjective workload ratings between the four segments; however, the blink frequencies did not show any significant difference in the workload between segments. The subjective and physiological workload data also partially supported the hypothesis that novice drivers would experience a higher workload level. The MCH ratings indicated that novice drivers experienced a higher workload; however, blink frequency indicated that the experienced drivers experienced a higher workload level.

Initially, it was hypothesized that blink frequency and electroencephalography measures of mental workload would be correlated. The data used in this study did not allow for this correlation to be performed. The EEG data were ultimately only able to be analyzed in a qualitative manner. Additionally, evidence was not found in this study to show that these two metrics would be correlated. Changes in mental workload would be categorized by increased blink frequency and increased EEG alpha and beta wave activity during tasks that are more mentally demanding. The blink frequency results were inconclusive regarding workload differences between the four course segments. The EEG data could not

be quantified to determine changes in workload levels between the four segments. This research does show that visual inspection of EEG alpha and beta wave activity can show changes in mental workload.

In addition to evaluating differences in workload assessment techniques, this research hypothesized that the miss rate and false alarm rate for IED detection would increase as workload increased. The findings of this research do not entirely support this hypothesis. While an increase in the miss rate and false alarm rate were not seen between the four segments, there was a difference in the miss rate and false alarm rate between experience levels. In general, the novices had significantly lower miss rates but significantly higher false alarm rates than the experienced drivers. A 0.08 average miss rate increase and a 0.09 average false alarm rate decrease were observed between the novice drivers and experienced drivers. These results make sense because the novices are more recently out of training. During training, they are instructed to call any item a potential IED regardless of their personal feeling or previous experience. Consequently, they would miss fewer items, but also increase the likelihood of identifying non-threats as potential ones.

The experiment focused on collecting physiological data through the use of EEG and eye tracking which can be highly correlated with mission performance, and subjective data with the Modified Cooper Harper Rating scale which indicates how difficult participants perceive their workload to be. When conducting testing in a real-world environment, it is important to utilize metrics that are pertinent, repeatable, and have a high likelihood of being utilized outside of a controlled environment. This research has provided some insight into the benefits and limitations of different technologies and will be useful in determining the most appropriate design for an operational setting. These results indicate that there is no single metric that provides a comprehensive picture of mental workload. Future research should utilize multiple measures in order to provide a complete analysis and understanding of the true workload faced in any particular situation.

Recommendations

The findings of this thesis support the need for additional research to better understand lapses in attention and degraded performance from many factors including lost situational awareness, mental workload, or other sources of workload and task demands. Minimally intrusive EEG, eye tracking, and other physiological metric equipment will aid in the data collection.

This study supports previous research indicating that novice drivers have degraded IED detection performance compared to more experienced drivers. Similarly, if drivers are not as skilled at assessing their surroundings, this could lead to increased instances of vehicle rollovers.

One limitation of this study was the size of the sample group. The study for this research use a sample size of 14 participants, but a larger sample size would have increased the robustness of the data. Additionally, a larger sample size would have allowed some substitution for participants who had more complete data sets for those who had poor quality data.

There are situations where EEG and eye blink data analysis are advisable assuming the technical difficulties are able to be overcome. Continuous metrics such as EEG or eye blink are useful whenever a continuous understanding of the changes in workload is required. Since the data can be collected consistently over time, these metrics are able to pinpoint the exact times or events where changes in workload take place. The downside to subjective ratings scales such as MCH is they only provide a snapshot. Even though individuals are asked to rate over a period of time, they are going to be more influenced by what they most recently experienced. Additionally, physiological metrics should be used to compare quantitative results across participants. Ordinal results generally do not produce meaningful results across participants since individuals interpret scaled data differently.

Different physiological metrics could be used to evaluate mental workload during a driving and detection task. Utilizing tools such as pupil diameter, heart rate variability, and skin conductance (Reimer and Mehler, 2011) that have been shown to measure mental workload would have allowed for more comparisons between metrics. Alternatively, more accurate tools for the metrics currently selected could

also be used. For example, a head-mounted eye tracking device would have eliminated many of the challenges that were encountered with the monitor-mounted eye tracker.

Future simulated data collection tasks should be more consistent between participants. The communications were not the same between all 14 participants because the original scripted communications may have questioned participants about items that an individual did not identify during their course. Therefore, the scripted communications were not able to be used. To enhance future studies, the course and communications should be setup in a way so that all participants have the same communications. Perhaps, the communications should be scripted in such a way that they are not based off of the items that participants are expected to see, but instead are similar to common vehicle chatter.

There are many other factors that have not been included in this research that are likely to impact an individual's mental workload while driving. These include factors such as stress, fatigue, duration of the drive, driving demands, task complexity, and overall vehicle control (de Waard, 1996). Future research is necessary to determine any significant impact these other factors may have on mental workload. Other dependent variables could also be included to study possible differences between novice and experienced drivers.

References

Backs, R. W., Lenneman, J. K., Wetzel, J. M., and Green, P. (2003). Cardiac Measures of Driver Workload during Simulated Driving with and without Visual Occlusion. *Human Factors, 45*(4), 525-538.

Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., and Montanari, R. (2011). Driver workload and eye blink duration. *Transportation Research Part F, 14*, 199-208.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., and Olmstead, R. (2004). Real-time analysis of EEG indices of alertness, cognition, and memory with a wireless EEG headset. *International Journal of Human Computer Interaction*, 151-170.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., and Craven, P. L. (2007). EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. *Aviation, Space, and Environmental Medicine*, *78*(5), B231-B244.

Berka, C., Levendowski, D. J., Ramsey, C. K., Davis, G., Lumicao, M. N., Stanney, K., Reeves, L., Harkness Regli, S., Tremoulet, P. D., and Stibler, K. (2005). Evaluation of an EEG-Workload Model in an Aegis Simulation Environment. Proceedings of SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations, Orlando, FL, SPIE: The International Society for Optical Engineering.

Biron, L. (2012). *Simulations Stress Operators to Reduce Stress Later*. Retrieved July 24, 2012, from Defense News website: http://www.defensenews.com/article/20120719/TSJ01/307190006/Simulations-Stress-Operators-Reduce-Stress-Later?odyssey=mod|newswell|text|FRONTPAGE|s

Brookhuis, K. A., and de Waard, D. (2001). Assessment of Drivers' Workload: Performance and Subjective and Physiological Indexes. In P. A. Hancock & P. A. Desmond, *Stress, Workload and Fatigue* (pp. 321-333). Laurence Erlbaum Associates Inc.

Brook, T. V. (2012). *IED attacks in Afghanistan set record*. Retrieved July 24, 2012, from Marine Corps Times website: http://www.marinecorpstimes.com/news/2012/01/gannett-ied-attacks-in-afghanistan-set-record-012512

Cain, B. (2007). A Review of the Mental Workload Literature. Defence Research and Development Toronto (Canada) (pp. 1-35). http://www.dtic.mil/cgi-bin/ GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA474193

Carsten, O., and Brookhuis, K. (2005). Issues arising from the HASTE experiments. *Transportation Research,* Part F(8), 191-196.

Chen, S., Epps, J., and Chen, F. (2011). A Comparison of Four Methods for Cognitive Load Measurement, *23rd Australian Computer-Human Interaction Conference*, 76-79.

Coles, M. G. H., and Rugg, M. D. (1995). Event-related brain potentials: an introduction. In Electrophysiology of mind: Event-related brain potentials and cognition (pp. 1-26). New York: Oxford University Press.

Collet, C., Petit, C., Champely, S., and Dittmar, A. (2003). Assessing Workload through Physiological Measurements in Bus Drivers Using an Automated System during Docking. Human Factors, 45(4), 539-548.

Congleton, J. J., Jones, W. A., Shiflett, S. G., McSweeney, K. P., and Huchingson, R. D. (1997). An Evaluation of Voice Stress Analysis Techniques in a Simulated AWACS Environment. *International Journal of Speech Technology,* 2, 61-69.
Craig, A. (1985). Signal Detection Theory and Probability Matching Apply to Vigilance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 29*(9), 9-13.

Cummings, M. L., and Nehme, C. E. (2009). *Modeling the Impact of Workload in Network Centric Supervisory Control Settings*. Massachusetts Institute of Technology.

de Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. The Netherlands: The Traffic Research Centre, University of Griningen.

Di Nocera, F., Camilli, M., and Terenzi, M. (2006). *Using the Distribution of Eye Fixations to Assess Pilots' Mental Workload*. Italy: Human Factors and Ergonomics Society.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., and Damos, D. L. (1991). Workload assessment in multi-task environments. In D. L. Damos (Ed.), *Mutliple-Task Performance* (pp. 207-216). http://books.google.com/books?hl=en&lr=&id=_KfaN4gKP8MC&oi=fnd&pg=PA207&dq=Workload+assessment+in+multi-task+environments&ots=S_B_Mh31ps&sig=QpyzihBk_z5saX4NpZx-hJw_80s#v=onepage&q=Workload%20assessment%20in%20multi-task%20environments&f=false

Embrey, D., Blackett, C., Marsden, P., and Peachey, J. (2006). *Development of a Human Mental Workload Assessment Tool.* Dalton Lancashire: Human Reliability Associates.

Engström, J., Johansson, E., and Östlud, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research*, Part F(8), 97-120.

Fukuda, K., Stern, J. A., Brown, T. B., and Russo, M. B. (2005). Cognition, Blinks, Eye-Movements, and Pupillary Movements During Performance of a Running Memory Task. *Aviaton, Space, and Environmental Medicine, 76*(7), C75-C85.

Gartner, W. B., and Murphy, M. R. (1976). *Pilot Workload and Fatigue: A Critical Survey of Concepts and Assessment Techniques*. Ames Research Center. <http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19770004731_1977004731.pdf>

Gattie, G., Mead, P., Baron, J., Hamilton, M., Iden, R., and Thurber, J. (2011). *Littoral Combat Ship Manpower, Workload, and Study Phase III Test Report: USS FREEDOM Rough Water Trials.* Dahlgren, VA: Naval Surface Warfare Center Dahlgren Division.

Gawron, V. J. (2000). *Human Performance Measures Handbook*. New Jersey: Lawrence Erlbaum Associates.

Gevins, A., and Smith, M. E. (2000). Neurophysiological Measure of Working Memory and Indvidual Differences in Cognitive Ability and Cognitive Style. *Cerebral Cortex, 10,* 829-839.

Gevins, A., Smith, M.E., McEvoy, L., and Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, 7, 374-385.

Gevins, A., Smith, M.E., Leong, H., et al. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors*, 40,79-91.

Hancock, P. A., and Meshkati, N. (1988). *Human Mental Workload*. The Netherlands: Elsevier Science Publishers.

Huey, B. M., and Wickens, C. D. (1993). *Workload Transition: Implications for Individual and Team Performance*. Washington, D.C.: National Academy Press.

Jamal, M. (2007). Job Stress and Job Performance Controversy Revisited: An Empirical Examination in Two Countries. *International Journal of Stress Management, 14*(2), 175-187.

Knoblauch, K., and Maloney, L. T. (2012). Signal Detection Theory. In *Modeling Psychophysical Data in R* (pp. 61-105). New York: Springer Science and Business Media.

Kohlmorgen, J., Dornhege, G., Braun, M. L., Blankertz, B., Müller, K. R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., and Kincses, W. E. (2007). Improving human performance in a real operating environment through real-time mental workload detection. *Toward Brain-Computer Interfacing*, 409-422.

Kramer, A. F., Trejo, L. J., and Humphrey, D. Assessment of mental workload with task-irrelevant auditory probes. Biological Psychology 1995, 40, 83-100.

Lee, J.D. (2000). A summary of the practical issues associated with mental workload measurement. Performance Analysis and Cognitive Technologies (PACT).

Lei, S., and Roetting, M. (2011). Influence of Task Contribution on EEG Spectrum Modulation for Driver Workload Estimation. *Human Factors, 53*(2), 168-179.

Leipold, J. D. (2009). *Study aims to identify IED detection experts*. Lanham, MD: Federal Information & News Dispatch, Inc.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A.L., Bittner, A. C., Jr., and Wherry, R.J. (1989). *Operator Workload: Comprehensive Review and Evaluation of Operator Workload Methodologies*. United States Army Research Institute for the Behavioral and Social Sciences.

Makeig, S. (1993). Auditory Event-Related Dynamics of the EEG Spectrum and Effects of Exposure to Tones. *Electroencephalography and Clinical Neurophysiology, 86*, 283-293. Retrieved from http://sccn.ucsd.edu/~scott/ersp93.html

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., and Brannan, J. R. (1990). Eye Movement Indices of Mental Workload. *Acta Psychologica, 75*(1), 75-89.

Megaw, E.D. (2005). The definition and measurement of mental workload. In J.R. Wilson and E.N. Corbett (Eds), *Evaluation of Human Work* (Third Edition). Boca Raton, FL: CRC Press.

Miller, M. W., Rietschel, J. C., McDonald, C. G., and Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. International Journal of Psychophysiology. 80, 75-78.

Miller, S. (2001). Workload Measures. Iowa City, IA: National Advanced Driving Simulator.
MIL-STD-1474D. (1997). *Depart of Defense Design Criteria Standard - Noise Limits* (AMSC A7245).

Murata, A. (2005). An Attempt to Evaluate Mental Workload Using Wavelet Transform of EEG. Human Factors, 47(5), 498-508

National Research Council. (1997). *Tactical display for soldiers: human factors considerations.* Washington, D.C.: National Academy Press.

NATO Research and Technology Organisation. (2001). *NATO Guidelines on Human Engineering Testing and Evaluation*. http://ftp.rta.nato.int/public//PubFullText/RTO/TR/RTO-TR-021///TR-021-03.pdf

Neill, D. (2011). Nursing Workload and the Changing Health Care Environment: A Review of the Literature. *Administrative Issues Journal: Education, Practice, and Research, 1*(2), 132-143.

Nomura, S. (2008). *Estimation of the Workload with Visual Display Terminal by Salivary Immune Substance*. Japan: IEEE Conference on Soft Computing in Industrial Applications.

Nuwer, M. R., Aminoff, M., Desmedt, J., Eisen, A. A., Goodin, D., Matsuoka, S., Mauguière, F., Shibasaki, H., Sutherling, W., and Vibert, J. (1994). IFCN recommended standards for short latency somatosensory evoked potentials. Report of an IFCN committee. *Electroencephalography and clinical Neurophysiology, 91*, 6-11.

Onton, J., Delorme, A., and Makeig, S. (2005). *Frontal midline EEG dynamics during working memory*. La Jolla, CA: Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego.

Perala, C. H., and Sterling, B. S. (2007). *Galvanic Skin Response as a Measure of Soldier Stress*. Maryland: U.S. Army Research Laboratory.

Pomplun, M., Sunkara, S., Fairley, A. V., and Xiao, M. (n.d.). Using Pupil Size as a Measure of Mental Workload in Video-Based Eye-Tracking. Department of Computer Science, University of Massachusetts Boston. <http://www.cs.umb.edu/~marc/pubs/pomplun_sunkara_fairley_xiao_draft.pdf>

Reimer, B. and Mehler, B. (2011). The impact of mental workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics*, 54(10), 932-942

Reimer, B., Mehler, B., Wang, Y., and Coughlin, J. F. (2012). A Field Study on the Impact of Variations in Short-Term Memory Demands on Drivers' Visual Attention and Driving Performance Across Three Age Groups. *Human Factors, 54*(3), 454-468.

Rubio, S., Diaz, E., Martin, J., and Puente, J. M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology: An International Review*, 53(1), 61-86.

Sanborn, J. K. (2010). *Rollover alert sought for use on MRAPs*. Retrieved July 23, 2012, from Marine Corps Times website: http://www.marinecorpstimes.com/news/2010/10/marine-mraps-rollover-100510w/

Santoro, T. P., Kieras, D. E., and Pharmer, J. A. (2004). *Verification and Validation of Latency and Workload Predictions for a Team of Humans by a Team of Computational Models*.

Shachtman, N. (2009). *The Army's New Land Warrior Gear: Why Soldiers Don't Like It*. Retrieved August 9, 2012. Popular Mechanics. <http://www.popularmechanics.com/technology/military/4215715>

Stanton, N., Salmon, P., Walker, G., Baber, C., and Jenkins, D. (2006). *Human Factors Methods: A Practical Guide for Engineering and Design.* Hampshire, England: Ashgate.

Tattersall, A. J., and Hockey, G. R. J. (1995). Level of Operator Control and Changes in Heart Rate Variability during Simulated Flight Maintenance. *Human Factors, 37*(4), 682-698.

Tsang, P. S., and Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics, 39*(3), 358-381.

U.S. General Accountability Office (2003). *Navy Actions Needed to Optimize Ship Crew Size and Reduce Total Ownership Costs.* Report number GAO-03-520. Washington, DC: GAO.

U.S. General Accountability Office (2010). *Littoral Combat Ship: Actions Needed to Improve Operating Cost Estimates and Mitigate Risks in Implementing New Concepts.* Report number GAO 10-257. Washington, DC: GAO.

Warm, J. S., Parasuraman, R., and Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors, 50,* 433-441.

Wickens, C. D., and Hollands, J. G. (2000). Engineering Psychology and Human Performance (Third Edition). New Jersey: Prentice Hall.

Wickens, C., Lee, J., Liu, Y., and Becker, S. (2004). An Introduction to Human Factors Engineering (Second Edition). New Jersey: Prentice Hall.

Wickens, T. D. (2002). Ideal observers and optimal performance. In Elementary Signal Detection Theory (pp. 32-35). New York: Oxford University Press.

Wierwille, W. W., and Casali, J. G. (1983). A Validated Rating Scale for Global Mental Workload Measurement Applications. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 27, 129-133.

Wierwille, W. W., and Eggemeier, F. T. (1993). Recommendations for Mental Workload Measurement in a Test and Evaluation Environment. *Human Factors,* 35(2), 263-281.

Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18,* 459-482.

Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M., and Atkins, M. S. (2012). Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surgical Endoscopy, 26,* 2746-2750.

Appendix A – Summary of Studies using Physiological Tools to Assess Mental Workload

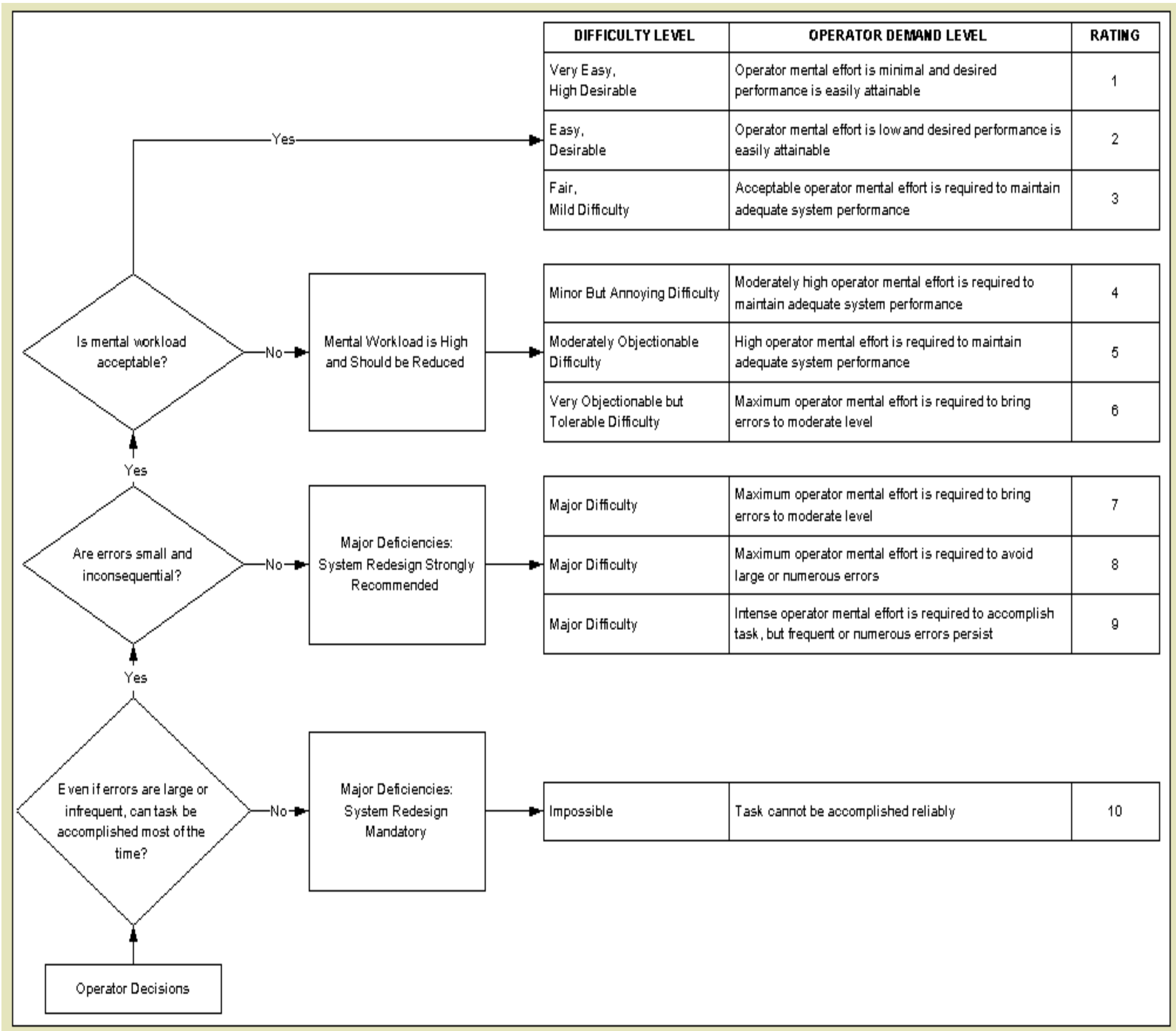
Physiological Tools		
Measurement Tool	***Environment Used***	***Summary of Findings***
Saccades (small eye movements)	Laboratory	Range of movements decreased and was severely restricted as workload increased. A decreased range was still observed even with task practice (May, Kennedy, Williams, Dunlap, and Brannan, 1990).
Saccades	Laboratory - Tone counting task	It was shown that as task difficulty increased, the number of spontaneous saccades decreased significantly. A limitation to these results is that they were obtained with an auditory task. There is the need to examine whether this same relationship exsits with a visual task (May, Kenneky, Williams, Dunlap, and Brannan, 1990).
Distribution of eye fixations	Laboratory - Flight simulator	During periods of high workload--take off and landing--eye fixations were more dispersed. Pilots would frequently move their fixation and the duration of fixation was very small. Results corresponded to NASA TLX scores (Di Nocera, Camilli, and Terenzi, 2006).
Pupil dilation	Laboratory	Pupil diameter was been shown to widen when an individual experienced increased mental workload. This response differed from the response to illumination (Pomplun, Sunkara, Fairley, and Xiao, n.d.)
Horizontal gaze concentration	Operational - Volvo XC90 driving on Route 93 north of Boston	As an individual's gaze became more concentrated, it suggested an increase in mental demand. This tunneling effect appeared before cognitive capacity was saturated or a decreased in driving performance appeared. An individual's gaze was the most narrowed at the highest level of mental demand (Reimer, Mehler, Wang and Coughlin, 2012).
EEG Spectrum Modulation	Laboratory - Driving Simulator	When the working memory loading task for drivers was increased, an increase in frontal thetal power/activity (fro-theta) and a decrease in parietal alpha power/activity (par-alpha) were found. An increase in the workload for the driving task produced a significant decrease in the par-alpha but had no change in the fro-theta. These results would suggest "a task-dependent workload effect on the modulation of EEG activity" (Lei and Roetting, 2011). Again, a signficant interaction effect was found in the EEG activity between the two tasks.
EEG - Event Related Potential (ERP), frontal midline theta, and frontal/parietal alpha	Laboratory	Subjects had their cognitive ability first tested with the WAIS-R. ERP amplitude proved to be higher for more abled subjects. The frontal midline theta was also enhanced when working memory was higher. Additionally, high-ability subjects used more of their parietal region when alpha waves were compared between frontal and parietal regions (Gevins and Smith, 2000).
EEG - wavelet analysis at theta, alpha and beta frequency bands	Laboratory - Matching task	Wavelet analysis was able to extract the total power and appearance time for these three frequency bands. It was found that total powre and appearance time increased as the task difficulty increased. These two metrics allowed the analysts to identify specific levels of task difficulty, indicating it is a good predictor (Murata, 2005).
ECG - Heart rate	Laboratory - Driving Simulator	Visual demand and heart rate were correlated to better understand how heart rate changes with visual demand when navigating curves of different radii. It was found that as the need for visual information increased, heart rate decreased. Previous research has shown that heart rate decreases as an individual absorbs and processes perceptual information (Backs, Lenneman, Wetzel, and Green, 2003).
ECG - Heart rate and HRV	Laboratory - Aircraft Simulator for flight engineer	During the exercise, average heart rate was elevated and HRV was surpressed compared to the baseline values. Heart rate and HRV were further examined between sessions (supervisory monitoring, routine fault rectification and open-ended problem solving), as well as phases

		(takeoff and landing). For heart rate, no significant differences were found between sessions but the takeoff and landing phases were higher than level flight. The sessions had no significant effects on HRV (Tattersall and Hockey, 1995).
Heart Rate and Heart Rate Variability (HRV)	Laboratory - Driving Simulator	Heart rate tends to increase with elevated levels of mental workload. Increased heart rate variability indicates reduced mental workload. However, HRV is only able to assess differences in major tasks (Brookhuis and de Waard, 2001).
Heart Rate and HRV	Laboratory - Driving Simulator	Drivers were subjected to tasks that varied both driving task load and working memory load. The increase in heart rate under both loading conditions was found to be significant as the task difficulty increased. A significant interaciton effect between the two task loading conditions was also found. HRV was found to have a significant negative correlation with workload across both tasks. However, for HRV, no significant interaction effect was identified (Lei and Roetting, 2011).
Heart Rate and Skin Conductance	Operational - Highway driving	As cognitive workload increased, so did heart rate. The heart rate pattern also followed consistently with a similar driving simulator study performed previously. The patterns for skin conductance between this operational study and the simulator were again similar. Skin conductance showed an increase with higher levels of workload. These two metrics are sensitive to initial changes in task demands (Reimer and Mehler, 2011).
Galvanic Skin Response (GSR)	Operational - M1043 truck	GSR increased under conditions that were perceived as more stressful by participants (Perala and Sterling, 2007).
Salivary immune substance - Immunoglobulin A (IgA)	Laboratory - Visual Display Terminal (VDT) - Performing calculations on a laptop	It was found that IgA increased immediately following a task and would decrease during the break period. The average IgA concentration during the second round of calculations was higher than the first (Nomura, 2008).
Electrodermal acitivity - ohmic perturbation duration (OPD)	Operational - Private track and open urban road	Electrodermal activity was measured to assess bus drivers' mental workload with the implementation of an automated assisstive system. The results showed that minitoring the system lead to an increased mental workload for drivers than when they manually controlled the bus. A learning effect was also noticed and after this effect was overcome, workload was shown to decrease when monitoring the system. If the system malfunctioned and required the driver to take control, increased workload was also shown (Collet, Petit, Champely, and Dittmar, 2003).

Appendix B – Summary of Studies using Subjective Tools to Assess Mental Workload

Subjective Tools		
Measurement Tool	***Environment Used***	***Summary of Findings***
NASA-TLX	Laboratory - Flight Simulator	Scores were higher during periods of elevated workload during flight--take off and landing (Di Nocera, Camilli, and Terenzi, 2006).
NASA-TLX	Laboratory - Matching task	After each task was completed, individuals were required to assess their workload using the NASA TLX. Results showed that the total workload score showed a significant increase as the task difficulty increased (Murata, 2005).
NASA-TLX	Laboratory - Driving Simulator	Drivers were subjected to tasks that varied both driving task load and working memory load. As the difficulty in both situations increased, perceived workload also increased. There was also a significant interaction effect identified (Lei and Roetting, 2011).
NASA-TLX, Subjective Workload Assessment Technique (SWAT), and Workload Profile (WP)	Laboratory - Sternberg's Memory Searching Task, Tracking Task, and Dual Tasks	All tasks proved to be minimally intrusive. WP has higher diagnosticity as compared to NASA TLX or SWAT. It is also highly sensitive to different task manipulations. Several recommendations were provided from this study. If comparing the mental workload of multiple tasks, use WP. If predicting the performance of an individual, use NASA TLX. If analyzing cognitive or attention demands of a task, use WP or SWAT (Rubio, Diaz, Martin and Puente, 2004).
WP, Bedford Scale and Psychophysical Scale	Laboratory - Sternberg's Memory Searching Task, and Tracking Task	The overall WP ratings proved to be less sensitive to task demands than the unidimensional scales. However, the WP as a multidimensional scale was able to be a diagnostic tool even under different task conditions (Tsang and Velazquez, 1996).
Visual-analog subjective workload (VSW) scales based on those used in SWAT	Laboratory - Aircraft Simulator for flight engineer	These assessments were used to indirectly measure workload associated with different flight maintenance activities. Due to the high amount of correlation across the subsystems, it can only be concluded at a high level that performance was associated with subjective workload ratings. Therefore, the subsystems that were rated as more mentally demanding had more faults and required more problem-solving (Tattersall and Hockey, 1995).

Appendix C – Modified Cooper-Harper Scale



Appendix D – Consent Form

Consent Form

Army Research Laboratory, Human Research & Engineering Directorate
Aberdeen Proving Ground, MD 21005

Title of Project: **Dynamic Classification of Soldier State**

Project Number: ARL 10-051

Sponsor: This research is sponsored by the Army Research Laboratory.

Principal Investigator: Brent Lance, PhD, 410-278-5943, brent.j.lance@us.army.mil
ARL/HRED, RDRL-HRS-C, Aberdeen Proving Ground, MD 21005

You are being asked to participate in a research study. This consent form explains the research study and your part in it. Please read this form carefully before you decide to take part. You can take as much time as you need. Please ask the research staff any questions at any time about anything you do not understand. You are a volunteer. If you join the study, you can change your mind later. You can decide not to take part now or you can quit at any time without penalty or negative consequences to you.

Purpose of the Study

The purpose of the study is to investigate the interactions relation between brain signals and driver performance in simulated driving scenarios. We also want to know how your body reacts as you perform different tasks, so we may also put sensors on your body to record changes in breathing rate, skin temperature, eye movements, and/or brain waves, as you perform various tasks.

Procedures to be Followed

You will be participating in a research study that simulates driving scenarios. First, you will be given a safety briefing about the equipment and asked to complete a questionnaire that records minimal demographic data and that helps ensure your safety and quality of data to be captured while you perform tasks.

Your participation will occur over one day. First, you will complete a practice mission to become familiar with the tasks. Then, you will be asked to complete two operational missions on the simulator.

During the missions, we may use sensors to record changes in your body's response to the tasks, such as changes in breathing rate, changes in skin temperature, eye movements, and/or changes in your brain waves. To measure your breathing rate, a strap would be fitted around your chest. To measure skin changes, a sensor would be placed on either your hand or your foot. To measure eye movements, a camera-based system will track the movement of your pupil. To measure changes in your brain waves, a head piece which holds the sensors next to your head would be fitted to your scalp. These brain wave sensors require an electrode gel to help establish a connection between the sensors and your skin. A sterile blunt needle will be used apply the gel without puncturing the skin. The EEG equipment will be attached and monitored by a trained scientist at all times.

During the study, you will perform a series of tasks that require different types of responses. When fulfilling the role of Driver, you will be asked to drive the simulated vehicle using a driver crew station, and you will be able to interact with the experimenters through the audio communication system. The experimenters will be able to monitor you and the equipment throughout the experiment, and your audio and video will be recorded for analysis purposes only. You will be able to take breaks as needed. Any data collected will never be used in conjunction with your name or any identifiable information.

The time required to complete the study session will not exceed 8 hours of simulator time over the course of the experimental day.

Discomforts and Risks

Risks associated with this research study are minimal and are equivalent to those encountered when you ride in a vehicle over urban terrain with no traffic.

The simulator aims to replicate typical vehicle motion experienced by a combat or tactical vehicle traversing urban terrain, and the whole body vibration caused by the simulation is not considered injurious. There is a risk of motion sickness or sickness resulting from the use of the simulator, but you will be monitored for signs of sickness throughout the session. Signs of sickness include but are not limited to nausea, cold sweating, pallor, and vomiting. The experiment can be terminated at any time if signs of sickness occur.

If sensors are used to assess changes in your body, you will be asked to report any discomfort arising from the sensors as they are applied and during the session. All systems are passive recording devices, and they have no known risks when used in this manner. If the EEG system is used today, special care will be taken to ensure that alcohol does not enter your eyes or ears when the experimenter is carefully cleaning the skin and applying electrode gel to connect the electrode to the skin. The electrode gel and adhesive tape used to secure sensors in place may

cause skin irritation for some participants, and eye irritation could be a risk if the electrode gel inadvertently comes into contact with your eyes. The investigator will wear polypropylene or nitrile gloves when cleaning your skin with rubbing alcohol. You could suffer an allergic reaction if you have allergies to rubbing alcohol touching the skin. If you are allergic to rubbing alcohol, you cannot participate in this experiment.

After each use, electrode caps are washed in hot soapy water and sterilized with a disinfectant solution immediately following the test session. Manufacturer specifications for the EEG amplifier confirm that risk of electrical shock is minimal or non-existent under the described conditions.

Benefits

There are no personal benefits to you for being in this study. However, the results of this study will provide information about how drivers and commanders interact in a simulated operational setting. This knowledge may lead to an improvement in the ability to assess Soldiers in complex, operational environments in order to enhance their performance.

Duration

Your participation will involve one session on one day. This session should take no more than 8 hours to complete. If sensors are applied to record changes in your body, the setup time will be no longer than 35-50 minutes.

Compensation for Participation

You cannot be paid if you are an active-duty member of the military or an employee of the Federal government.

Confidentiality

Your participation in this research is confidential. Each participant will be assigned a unique, non-personally identifying ID number that will be used on all questionnaires, data files, and data logs. The data collected today will be stored and secured in a locked room. Publication of the results of this study in a journal or technical report or presentation at a meeting will not reveal personally identifiable information. This consent form will be retained by the principal investigator for a minimum of three years.

The research staff will protect your data from disclosure to people not connected with the study. However, complete confidentiality cannot be guaranteed because officials of the U. S. Army Human Research Protections Office and the Army Research Laboratory's Institutional Review Board are permitted by law to inspect the records obtained in this study to ensure compliance with laws and regulations covering experiments using human subjects.

Contact Information for Additional Questions

You have the right to obtain answers to any questions you might have about this research both while you take part in the study and after you leave the research site. Please contact anyone listed at the top of the first page of this consent form for more information about this study. You may also contact the Chairperson of the Human Research & Engineering Directorate, Institution Review Board, at (410) 278-5992 with questions, complaints, or concerns about this research, or if you feel this study has harmed you. The Chairperson can also answer questions about your rights as a research participant. You may also call the Chairperson's number if you cannot reach the research team or wish to talk to someone else.

Voluntary Participation

Your decision to be in this research is voluntary. You can stop at any time. You do not have to answer any questions you do not want to answer. Refusal to take part in or withdrawal from this study will involve no penalty or loss of benefits you would receive by staying in it.

Military personnel cannot be punished under the Uniform Code of Military Justice for choosing not to take part in or withdrawing from this study, and cannot receive administrative sanctions for choosing not to participate.

Civilian or contractor personnel cannot receive administrative sanctions for choosing not to participate in or withdrawing from this study.

Principal Investigator: Brent Lance
Reviewed: 12 April 2012

ARMY RESEARCH LABORATORY
PROJECT NO. ARL 10-051
APPROVAL EXPIRES 19 September 2012

You must be between 18 and 45 years of age to take part in this research study. If you agree to take part in this research study based on the information outlined above, please sign your name and indicate the date below.

You will be given a copy of this consent form.

This consent form is approved from 20 September 2011to 19 September 2012

Do not sign this consent form after the expiration date of: 19 September 2012

Participant's Signature	Date
Participant's Printed Name	
Signature of Person Obtaining Consent	Date
Printed Name of Person Obtaining Consent	

Appendix E – Statistical Analysis Results

```
Two-sample T for Years of Service

Level  N   Mean  StDev  SE Mean
EXP    8   8.69   1.83     0.65
NOV    6  3.417  0.492     0.20


Difference = mu (EXP) - mu (NOV)
Estimate for difference:  5.271
95% CI for difference:  (3.708, 6.834)
T-Test of difference = 0 (vs not =): T-Value = 7.78  P-Value = 0.000
```

Figure 18: Minitab Two-Sample T-Test for Years of Service

```
Analysis of Variance for MCH Rating, using Adjusted SS for Tests

Source               DF   Seq SS  Adj SS  Adj MS      F      P
Subject(Experience)  12   81.582  74.371   6.198   5.21  0.000
Segment               3   55.715  60.944  20.315  17.08  0.000
Experience            1   61.744  62.966  62.966  52.95  0.000
Segment*Experience    3    2.281   2.281   0.760   0.64  0.596
Error                30   35.678  35.678   1.189
Total                49  237.000


S = 1.09054   R-Sq = 84.95%   R-Sq(adj) = 75.41%
```

Figure 19: ANOVA for Modified Cooper Harper Rating

```
                         Individual 95% CIs For Mean Based on
                         Pooled StDev
Level   N   Mean  StDev  ------+---------+---------+---------+---
1      14  3.179  1.908  (------*------)
2      13  4.346  1.784          (------*------)
3      12  5.458  1.777                 (------*-------)
4      11  6.227  2.240                      (-------*------)
                         ------+---------+---------+---------+---
                             3.0       4.5       6.0       7.5

Pooled StDev = 1.925


Grouping Information Using Tukey Method

Segment   N   Mean  Grouping
4        11  6.227  A
3        12  5.458  A
2        13  4.346  A B
1        14  3.179    B

Means that do not share a letter are significantly different.
```

Figure 20: Tukey Test between Segments for Modified Cooper Harper Rating

```
                          Individual 95% CIs For Mean Based on Pooled StDev
Level   N   Mean  StDev    +---------+---------+---------+---------
EXP    29  3.759  1.912    (-------*------)
NOV    21  6.000  1.917                          (-------*-------)
                           +---------+---------+---------+---------
                          3.0       4.0       5.0       6.0

Pooled StDev = 1.914


Grouping Information Using Tukey Method

Experience   N   Mean  Grouping
NOV         21  6.000  A
EXP         29  3.759    B

Means that do not share a letter are significantly different.
```

Figure 21: Tukey Test between Experience Levels for Modified Cooper Harper Rating

```
Analysis of Variance for MR, using Adjusted SS for Tests

Source                DF   Seq SS   Adj SS   Adj MS      F      P
Subject(Experience)   12  0.64192  0.57700  0.04808   2.39  0.023
Segment                3  1.85940  1.70551  0.56850  28.26  0.000
Experience             1  0.06482  0.06340  0.06340   3.15  0.085
Segment*Experience     3  0.03966  0.03966  0.01322   0.66  0.584
Error                 34  0.68391  0.68391  0.02011
Total                 53  3.28970


S = 0.141827   R-Sq = 79.21%   R-Sq(adj) = 67.59%
```

Figure 22: ANOVA for IED Miss Rate

```
                                Individual 95% CIs For Mean Based on
                                Pooled StDev
Level   N    Mean   StDev   ----+---------+---------+---------+-----
1      14  0.0000  0.0000   (---*---)
2      14  0.3800  0.1804                      (---*---)
3      13  0.3808  0.2267                     (----*----)
4      13  0.4985  0.1605                           (----*---)
                            ----+---------+---------+---------+-----
                              0.00      0.20      0.40      0.60

Pooled StDev = 0.1642


Grouping Information Using Tukey Method

Segment   N    Mean  Grouping
4        13  0.4985  A
3        13  0.3808  A
2        14  0.3800  A
1        14  0.0000    B

Means that do not share a letter are significantly different.
```

Figure 23: Tukey Test between Segments for IED Miss Rate

```
Analysis of Variance for MR w/o Segment 1, using Adjusted SS for Tests

Source              DF   Seq SS   Adj SS   Adj MS     F      P
Sub(Exp)            12  0.78054  0.80376  0.06698  3.22  0.008
Segments 2-4         2  0.10656  0.09691  0.04846  2.33  0.121
Exp                  1  0.12083  0.11120  0.11120  5.35  0.030
Segments 2-4*Exp     2  0.00614  0.00614  0.00307  0.15  0.863
Error               22  0.45716  0.45716  0.02078
Total               39  1.47124


S = 0.144152   R-Sq = 68.93%   R-Sq(adj) = 44.92%
```

Figure 24: ANOVA for IED Miss Rate with Segments 2-4

```
Grouping Information Using Tukey Method and 95.0% Confidence

Exp   N    Mean  Grouping
EXP  24  0.4513  A
NOV  16  0.3360    B

Means that do not share a letter are significantly different.
```

Figure 25: Tukey Test between Experience Levels for IED Miss Rate in Segments 2-4

Analysis of Variance for FAR, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject(Experience)	12	0.35943	0.39479	0.03290	1.24	0.299
Segment	3	0.15114	0.12622	0.04207	1.58	0.212
Experience	1	0.14897	0.16061	0.16061	6.04	0.019
Segment*Experience	3	0.15040	0.15040	0.05013	1.89	0.151
Error	34	0.90424	0.90424	0.02660		
Total	53	1.71417				

S = 0.163080   R-Sq = 47.25%   R-Sq(adj) = 17.77%

Figure 26: ANOVA for IED False Alarm Rate

```
                              Individual 95% CIs For Mean Based on
                              Pooled StDev
Level   N    Mean   StDev   -+---------+---------+---------+--------
EXP    32  0.1750  0.1826   (---------*----------)
NOV    22  0.2728  0.1626                 (-----------*------------)
                            -+---------+---------+---------+--------
                            0.120     0.180     0.240     0.300

Pooled StDev = 0.1748
```

Grouping Information Using Tukey Method

Experience	N	Mean	Grouping
NOV	22	0.2728	A
EXP	32	0.1750	B

Means that do not share a letter are significantly different.

Figure 27: Tukey Test between Experience Levels for IED False Alarm Rate

Analysis of Variance for Normalized Freq, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject(Experience)	12	5082.50	5014.17	417.85	28.96	0.000
Segment	3	16.36	15.35	5.12	0.35	0.786
Experience	1	75.19	79.14	79.14	5.49	0.026
Segment*Experience	3	8.49	8.49	2.83	0.20	0.898
Error	29	418.42	418.42	14.43		
Total	48	5600.95				

S = 3.79845   R-Sq = 92.53%   R-Sq(adj) = 87.64%

Figure 28: ANOVA for Blink Frequency

```
Grouping Information Using Tukey Method and 95.0% Confidence

Experience   N   Mean  Grouping
EXP         27  9.581  A
NOV         22  6.894    B

Means that do not share a letter are significantly different.
```

Figure 29: Tukey Test between Experience Levels for Blink Frequency

```
Two-sample T for Blink Rate (H/FA) vs Blink Rate (M/CR)

                    N  Mean  StDev  SE Mean
Blink Rate (H/FA)  40   9.3   10.0      1.6
Blink Rate (M/CR)  44   9.3   10.4      1.6


Difference = mu (Blink Rate (H/FA)) - mu (Blink Rate (M/CR))
Estimate for difference:  0.03
95% CI for difference:  (-4.41, 4.48)
T-Test of difference = 0 (vs not =): T-Value = 0.02  P-Value = 0.988  DF = 81
```

Figure 30: Minitab Two-Sample T-Test for 2 Blink Rate Categories

```
Analysis of Variance for Blink Duration, using Adjusted SS for Tests

Source                DF   Seq SS   Adj SS  Adj MS     F      P
Subject(Experience)   12  59363.5  53233.9  4436.2  4.49  0.000
Segment                3   1072.0   1295.6   431.9  0.44  0.728
Experience             1    316.7    182.0   182.0  0.18  0.671
Segment*Experience     3   1059.0   1059.0   353.0  0.36  0.784
Error                 29  28680.8  28680.8   989.0
Total                 48  90492.1


S = 31.4483   R-Sq = 68.31%   R-Sq(adj) = 47.54%
```

Figure 31: ANOVA for Blink Duration

```
Two-sample T for Blink Dur (H/FA) vs Blink Dur (M/CR)

                   N   Mean  StDev  SE Mean
Blink Dur (H/FA)  40  311.8   61.9      9.8
Blink Dur (M/CR)  44  312.5   69.8       11


Difference = mu (Blink Dur (H/FA)) - mu (Blink Dur (M/CR))
Estimate for difference:  -0.7
95% CI for difference:  (-29.3, 27.9)
T-Test of difference = 0 (vs not =): T-Value = -0.05  P-Value = 0.961  DF = 81
```

Figure 32: Minitab Two-Sample T-Test for 2 Blink Duration Categories

```
The regression equation is
MR = 0.105 + 0.0509 MCH Rating - 0.000453 Blink Duration
     + 0.00676 Normalized Freq


40 cases used, 16 cases contain missing values


Predictor              Coef    SE Coef      T      P
Constant             0.1048     0.2541   0.41  0.682
MCH Rating          0.05088    0.01622   3.14  0.003
Blink Duration   -0.0004532  0.0008102  -0.56  0.579
Normalized Freq    0.006763   0.003192   2.12  0.041


S = 0.218416   R-Sq = 25.5%   R-Sq(adj) = 19.3%
```

Figure 33: Minitab Regression Analysis for Miss Rate Explained by MCH Rating, Blink Duration and Blink Frequency

```
Analysis of Variance

Source          DF       SS       MS     F      P
Regression       3  0.58678  0.19559  4.10  0.013
Residual Error  36  1.71740  0.04771
Total           39  2.30418
```

Figure 34: ANOVA for Miss Rate Explained by MCH Rating, Blink Duration and Blink Frequency

```
The regression equation is
MR = 0.0251 + 0.0434 MCH Rating + 0.00653 Normalized Freq


45 cases used, 11 cases contain missing values


Predictor            Coef   SE Coef     T      P
Constant          0.02507   0.08847  0.28  0.778
MCH Rating        0.04341   0.01540  2.82  0.007
Normalized Freq  0.006533  0.003076  2.12  0.040


S = 0.224490   R-Sq = 20.6%   R-Sq(adj) = 16.9%
```

Figure 35: Minitab Regression Analysis for Miss Rate Explained by MCH Rating and Blink Frequency

```
Analysis of Variance

Source          DF       SS       MS     F      P
Regression       2  0.55067  0.27534  5.46  0.008
Residual Error  42  2.11663  0.05040
Total           44  2.66730
```

Figure 36: ANOVA for Miss Rate Explained by MCH Rating and Blink Frequency

```
The regression equation is
MR = 0.103 + 0.0417 MCH Rating


50 cases used, 6 cases contain missing values


Predictor       Coef  SE Coef     T      P
Constant     0.10255  0.07683  1.33  0.188
MCH Rating   0.04167  0.01483  2.81  0.007


S = 0.228351   R-Sq = 14.1%   R-Sq(adj) = 12.3%
```

Figure 37: Minitab Regression Analysis for Miss Rate Explained by MCH Rating

```
Analysis of Variance

Source          DF       SS       MS     F      P
Regression       1  0.41154  0.41154  7.89  0.007
Residual Error  48  2.50293  0.05214
Total           49  2.91447
```

Figure 38: ANOVA for Miss Rate Explained by MCH Rating

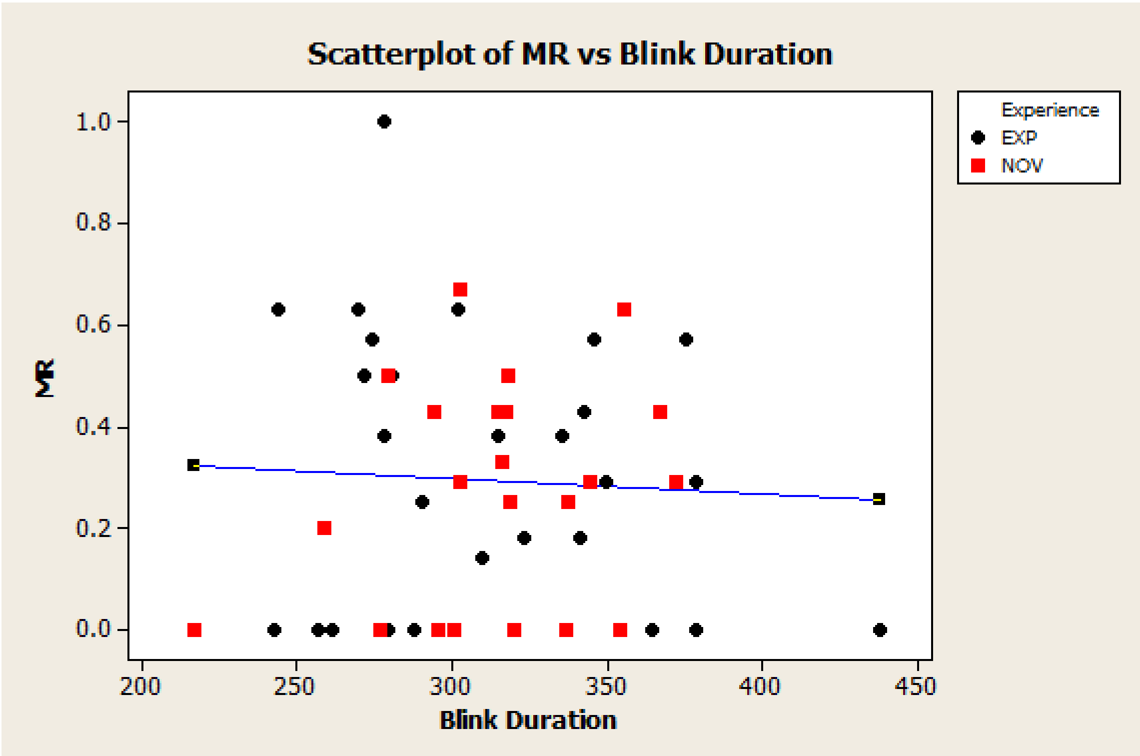


Figure 39: Scatterplot of Miss Rate versus Blink Duration

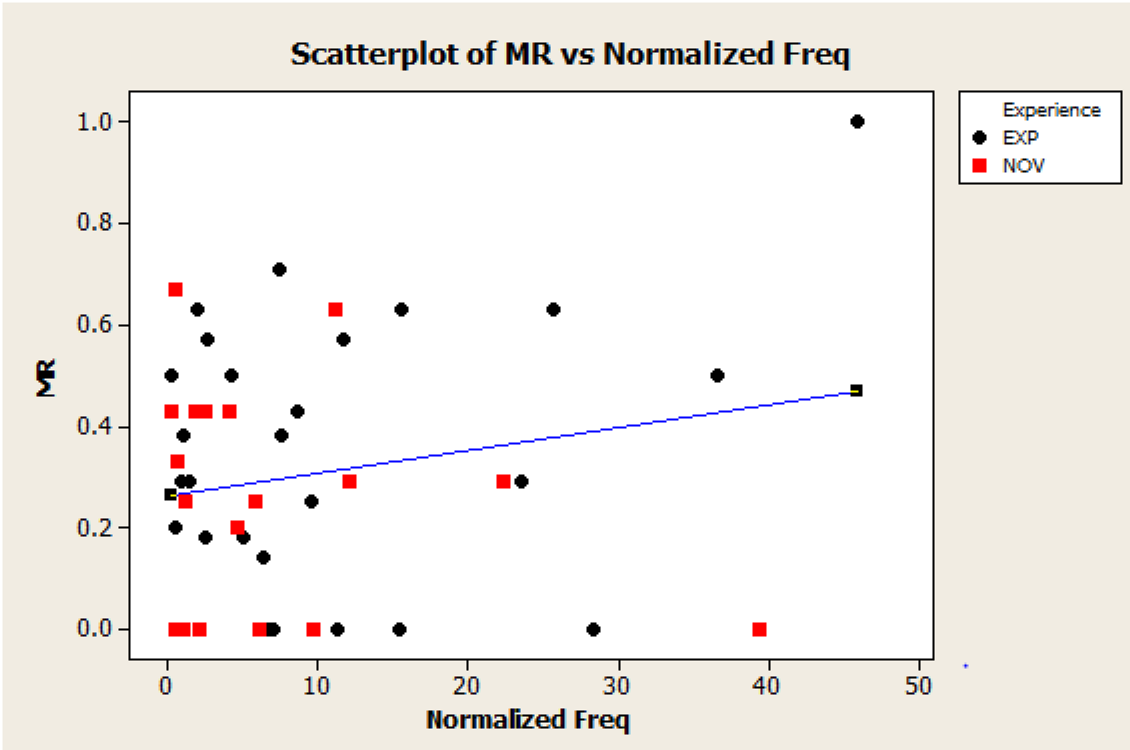


Figure 40: Scatterplot of Miss Rate versus Blink Frequency

Appendix F – Time-Frequency Analysis Graphs

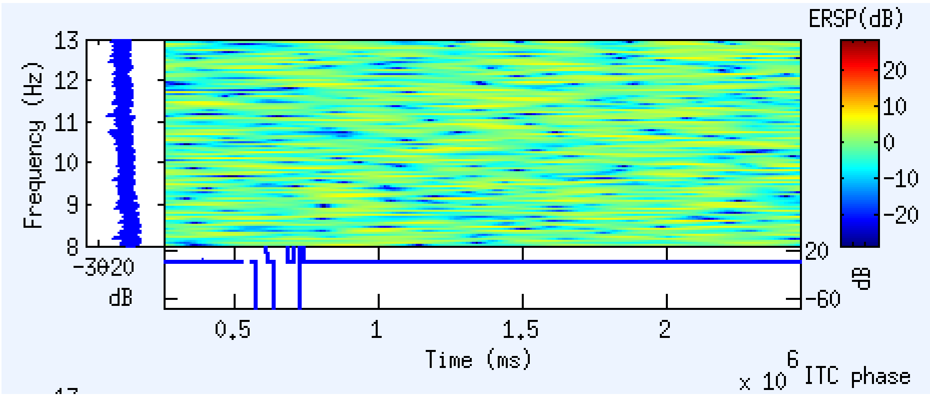


Figure 41: Participant 1 EEG Channel 4 Alpha Frequency

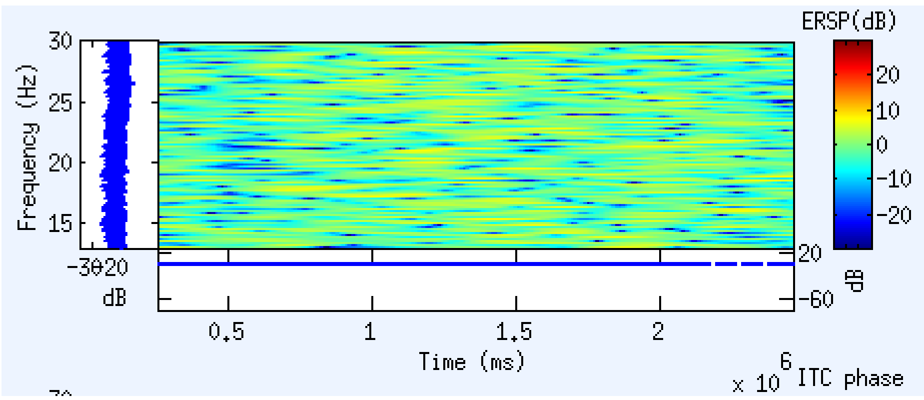


Figure 42: Participant 1 EEG Channel 4 Beta Frequency

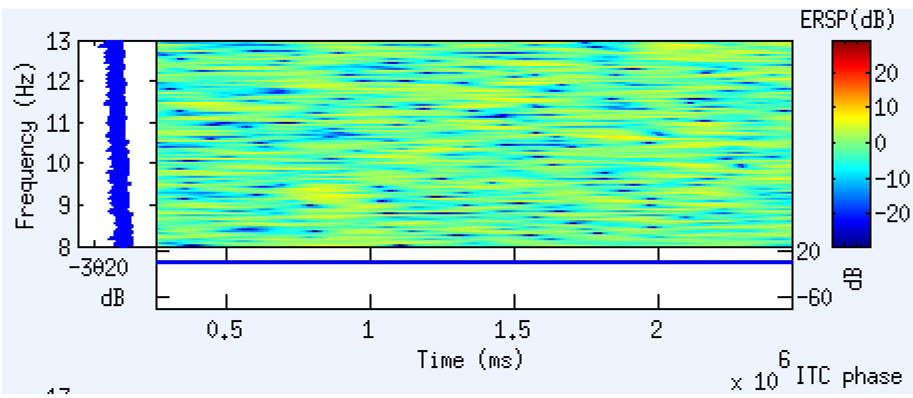


Figure 43: Participant 1 EEG Channel 11 Alpha Frequency

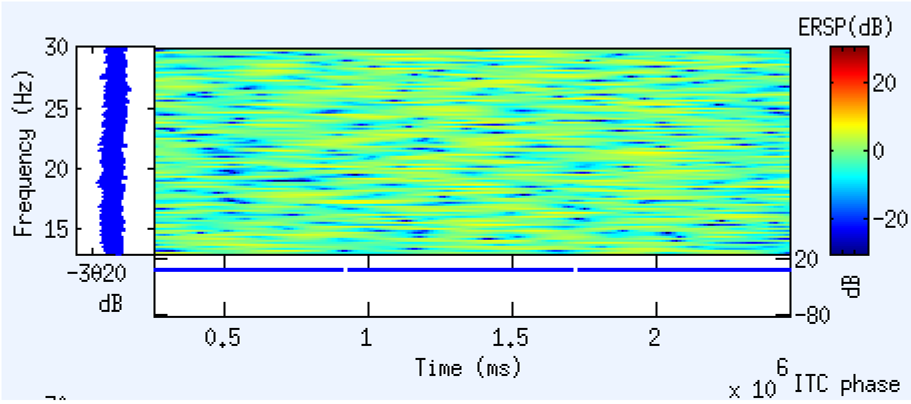


Figure 44: Participant 1 EEG Channel 11 Beta Frequency

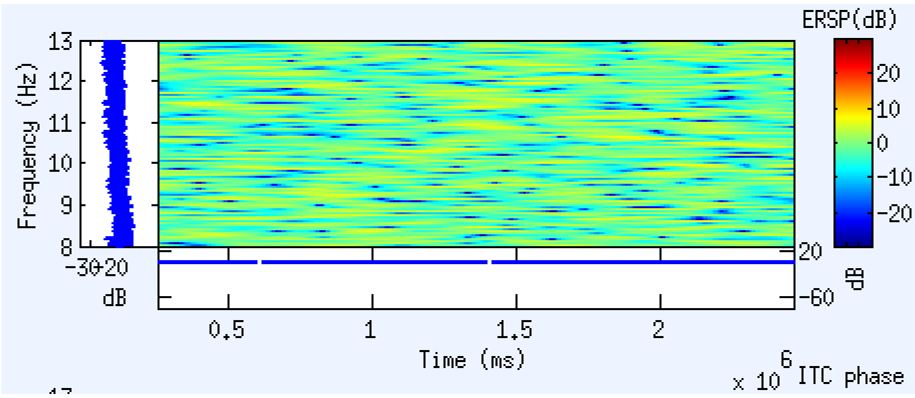


Figure 45: Participant 1 EEG Channel 38 Alpha Frequency

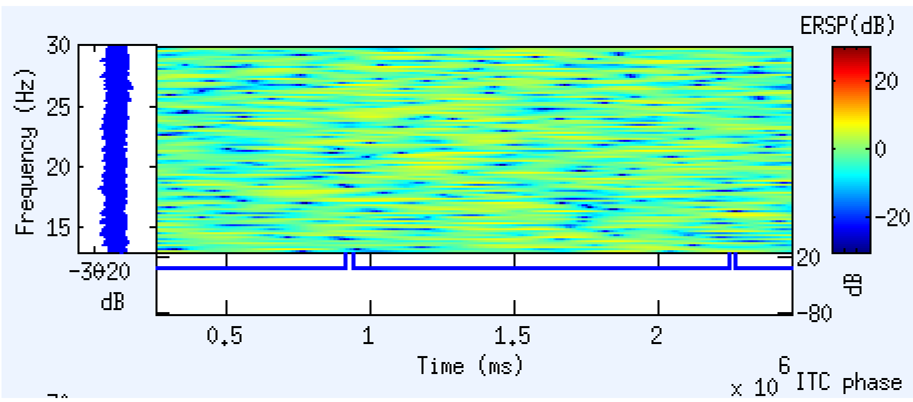


Figure 46: Participant 1 EEG Channel 38 Beta Frequency

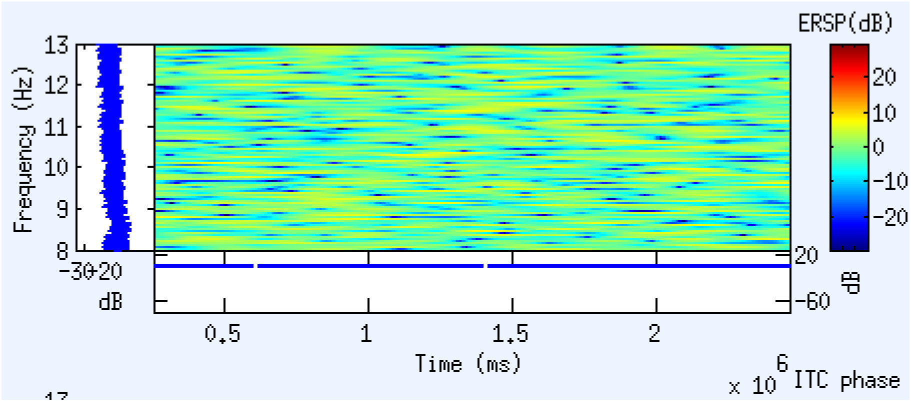


Figure 47: Participant 1 EEG Channel 39 Alpha Frequency

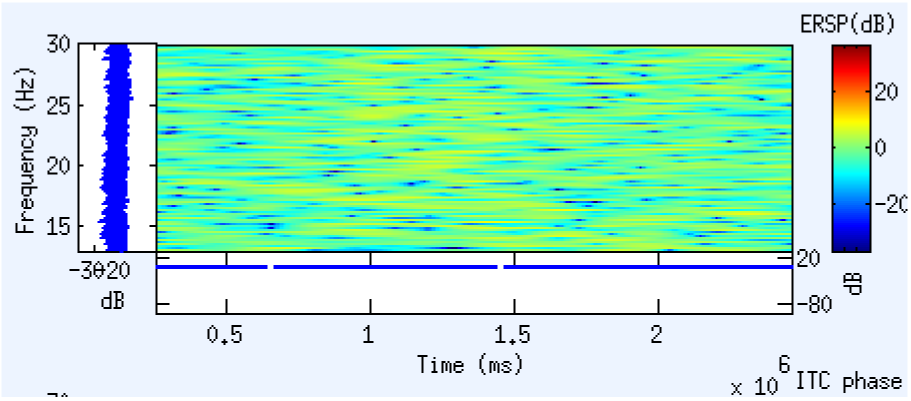


Figure 48: Participant 1 EEG Channel 39 Beta Frequency

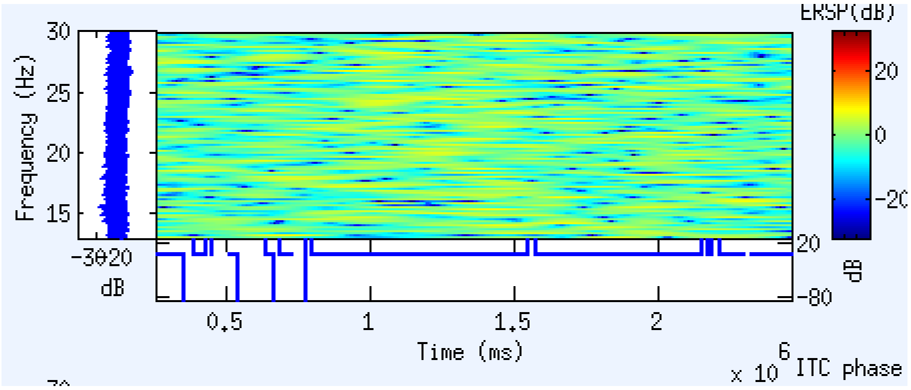


Figure 49: Participant 1 EEG Channel 46 Beta Frequency

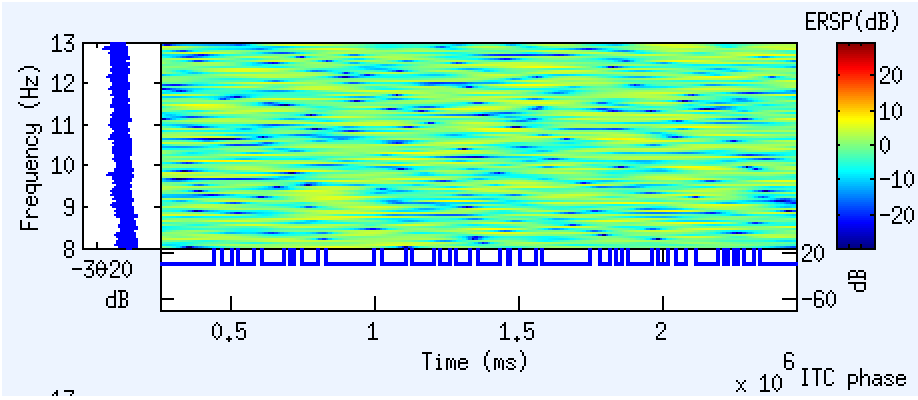


Figure 50: Participant 1 EEG Channel 47 Alpha Frequency

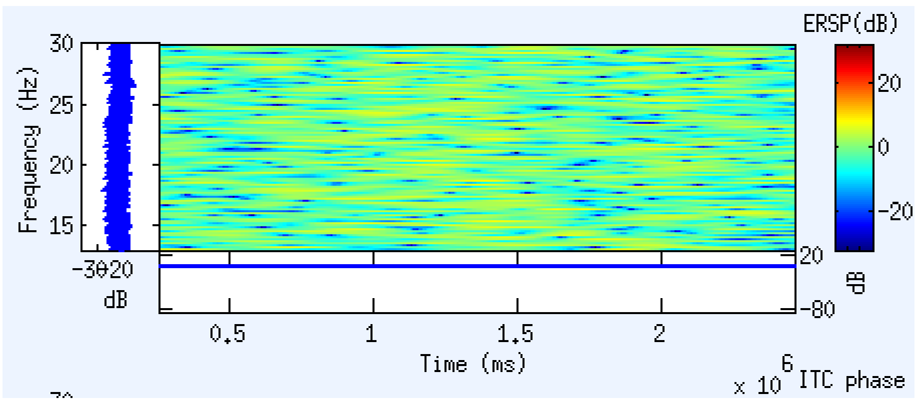


Figure 51: Participant 1 EEG Channel 47 Beta Frequency

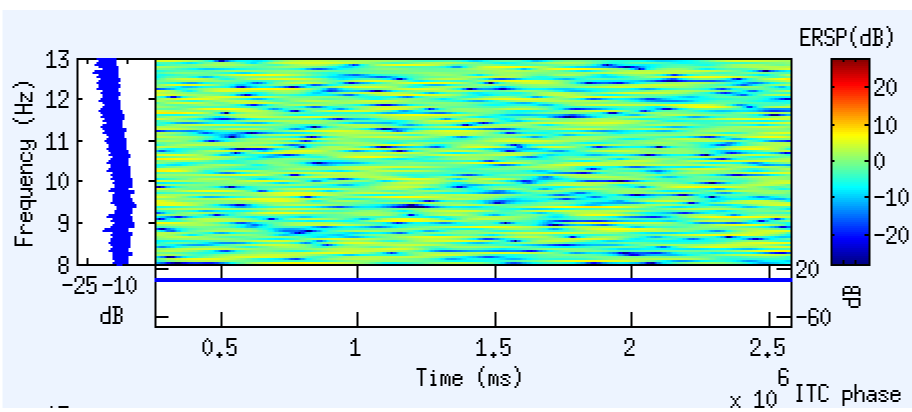


Figure 52: Participant 2 EEG Channel 4 Alpha Frequency

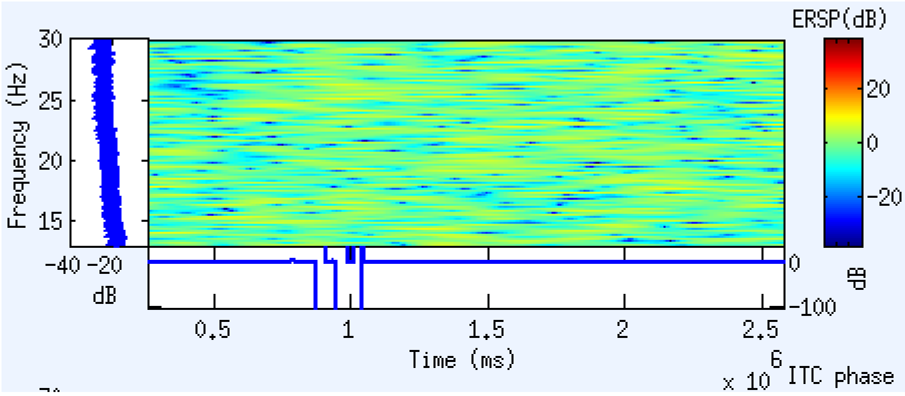


Figure 53: Participant 2 EEG Channel 4 Beta Frequency

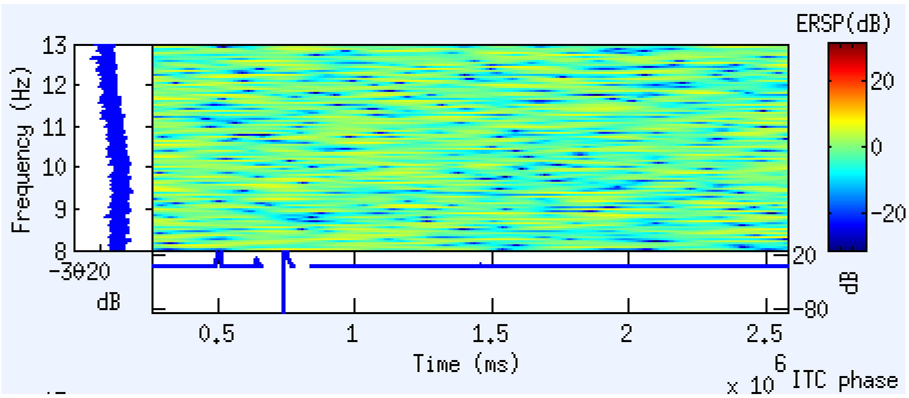


Figure 54: Participant 2 EEG Channel 11 Alpha Frequency

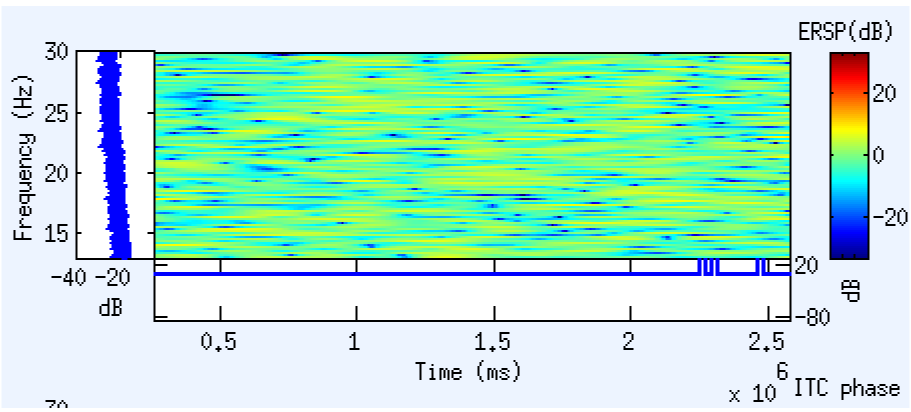


Figure 55: Participant 2 EEG Channel 11 Beta Frequency

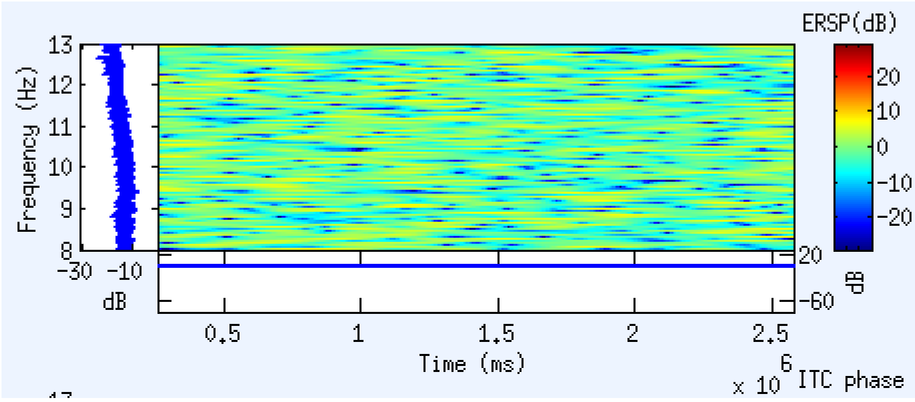


Figure 56: Participant 2 EEG Channel 38 Alpha Frequency

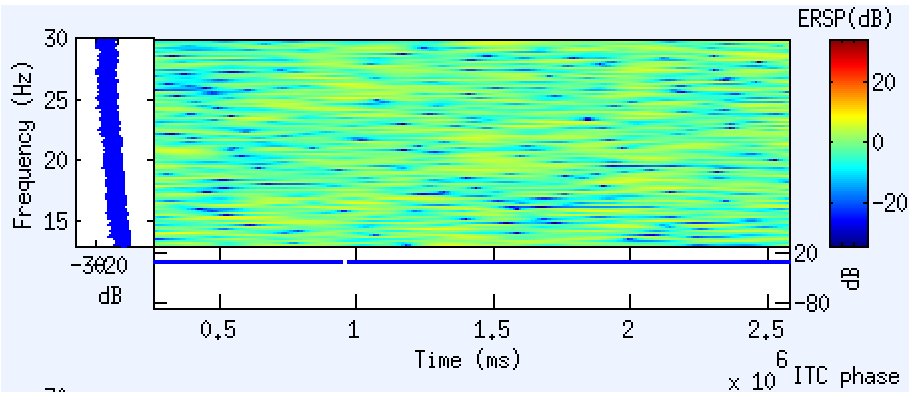


Figure 57: Participant 2 EEG Channel 38 Beta Frequency

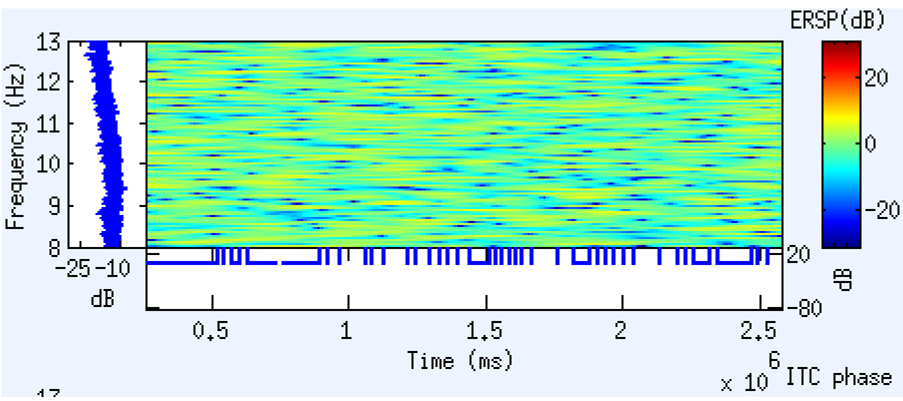


Figure 58: Participant 2 EEG Channel 39 Alpha Frequency

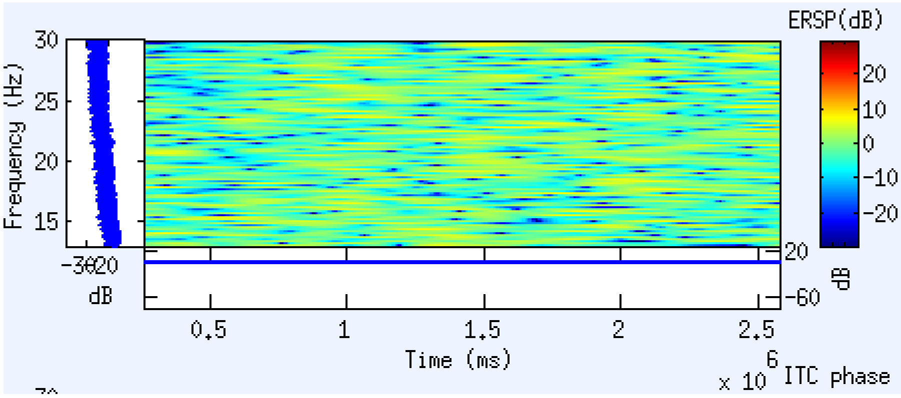


Figure 59: Participant 2 EEG Channel 39 Beta Frequency

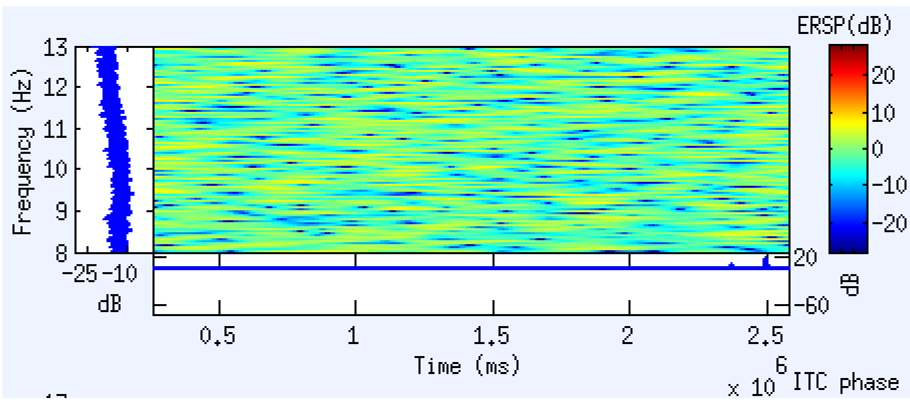


Figure 60: Participant 2 EEG Channel 46 Alpha Frequency

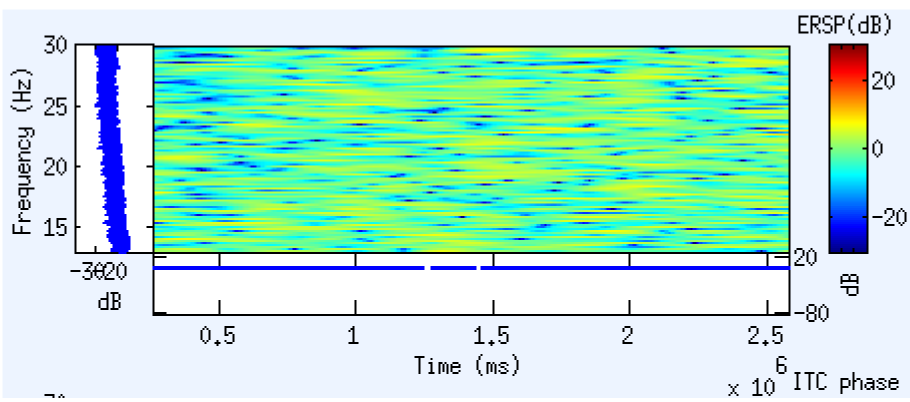


Figure 61: Participant 2 EEG Channel 46 Beta Frequency

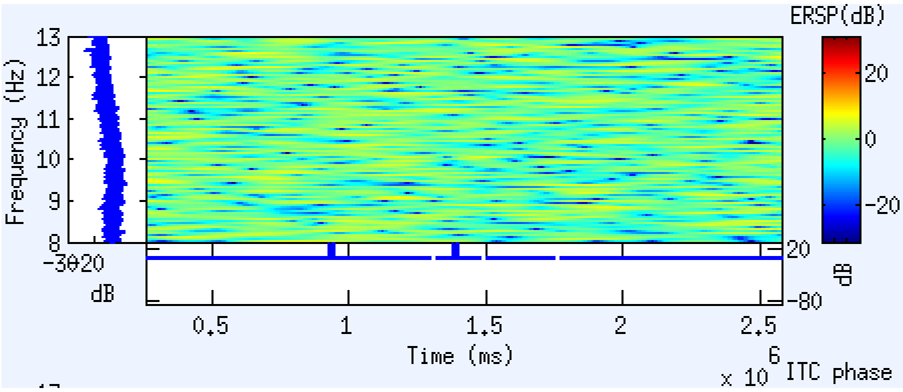


Figure 62: Participant 2 EEG Channel 47 Alpha Frequency

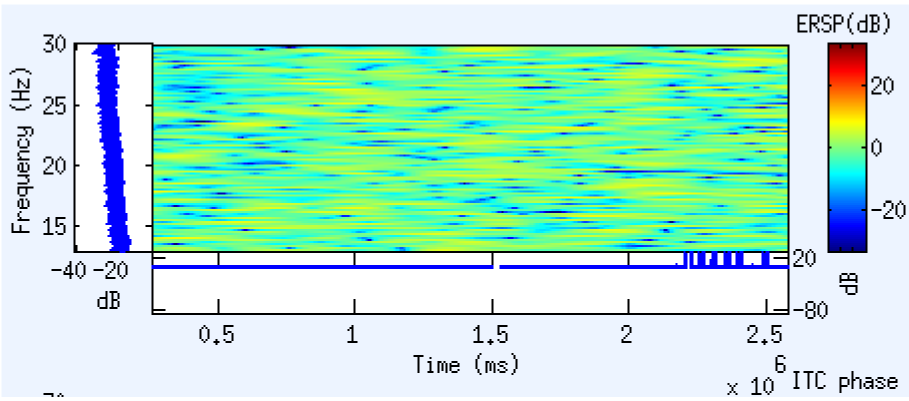


Figure 63: Participant 2 EEG Channel 47 Beta Frequency

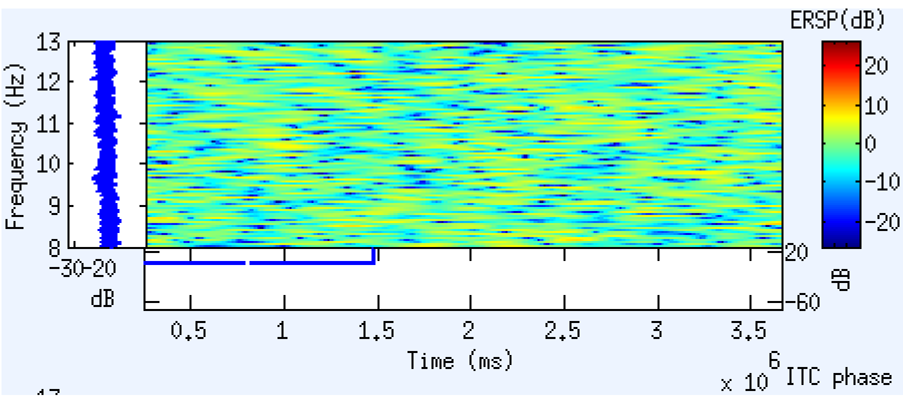


Figure 64: Participant 3 EEG Channel 4 Alpha Frequency

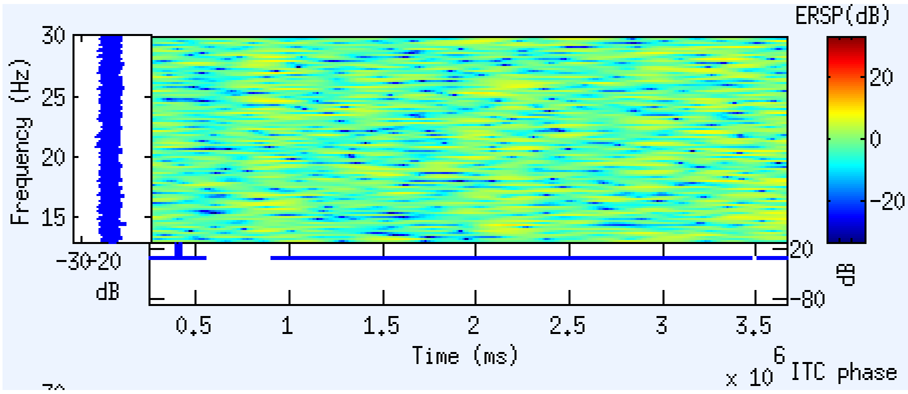


Figure 65: Participant 3 EEG Channel 4 Beta Frequency

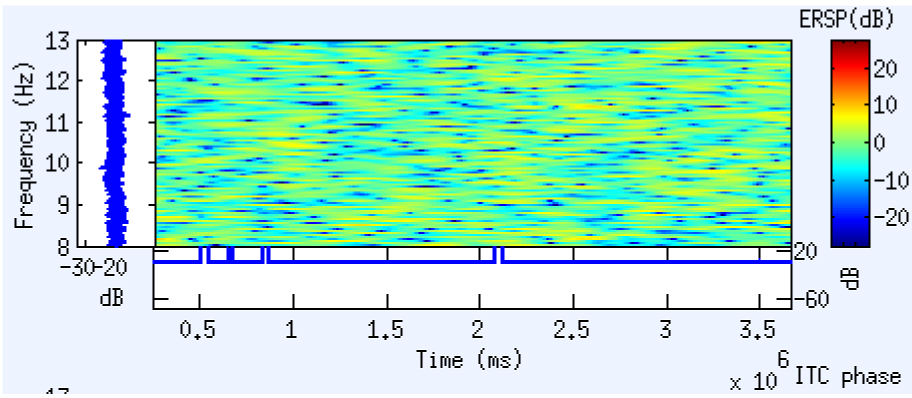


Figure 66: Participant 3 EEG Channel 11 Alpha Frequency

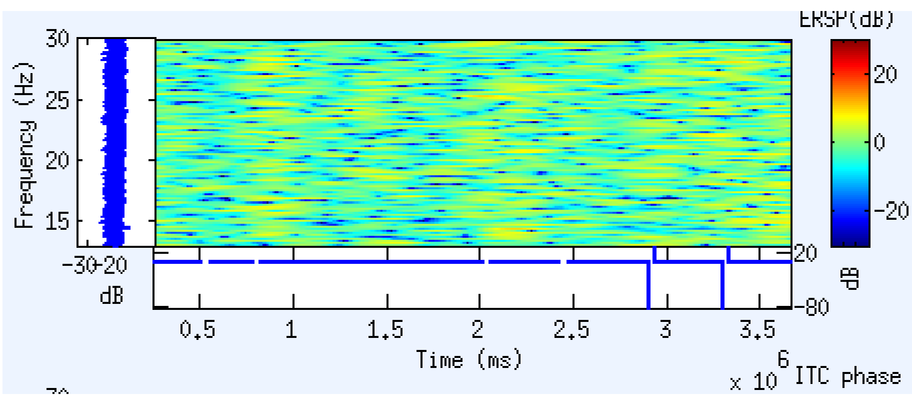


Figure 67: Participant 3 EEG Channel 11 Beta Frequency

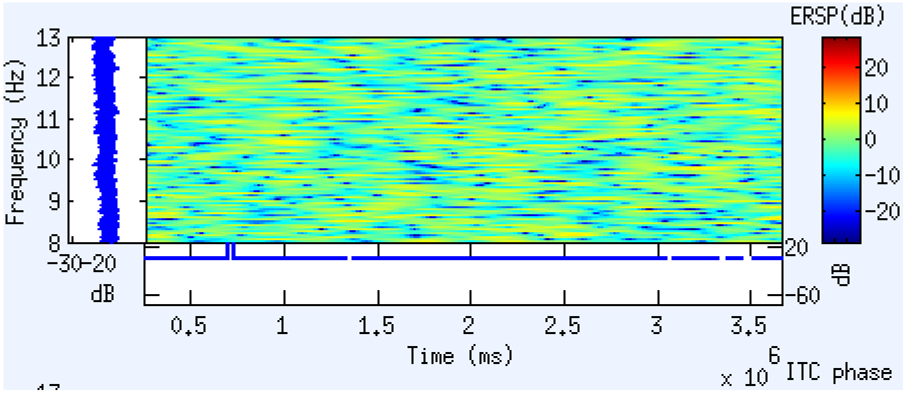


Figure 68: Participant 3 EEG Channel 38 Alpha Frequency

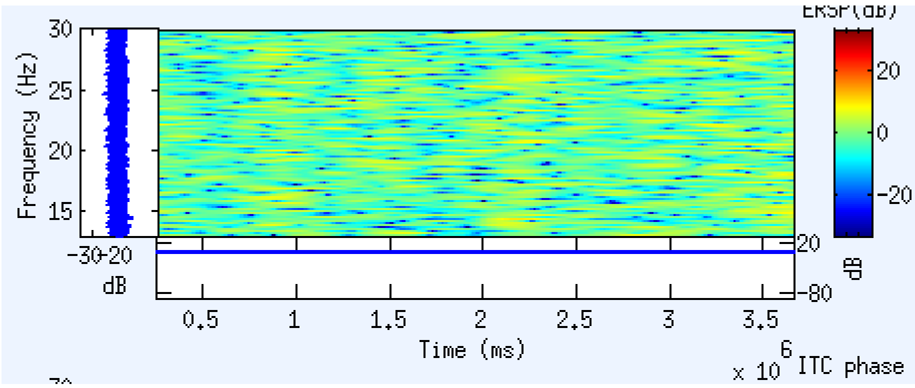


Figure 69: Participant 3 EEG Channel 38 Beta Frequency

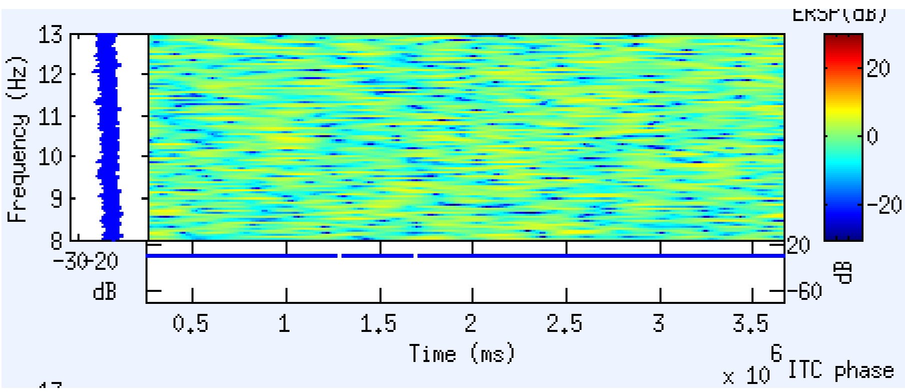


Figure 70: Participant 3 EEG Channel 39 Alpha Frequency

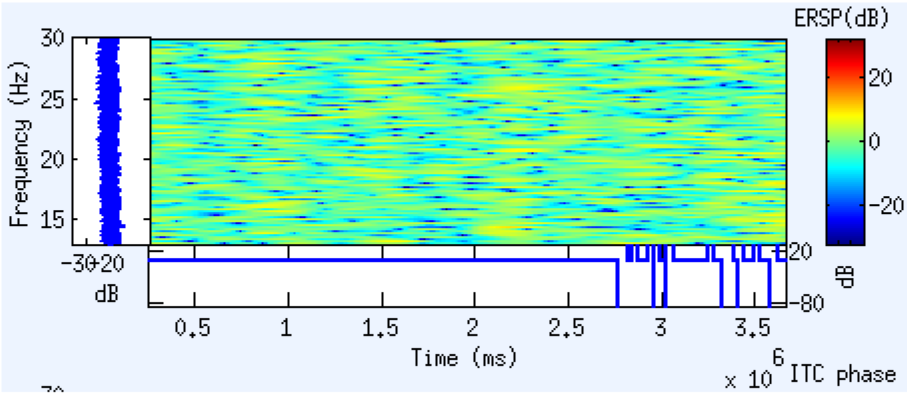


Figure 71: Participant 3 EEG Channel 39 Beta Frequency

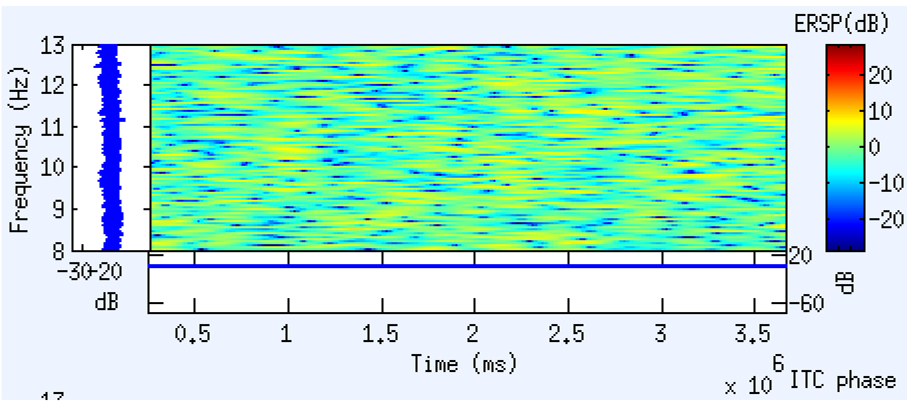


Figure 72: Participant 3 EEG Channel 46 Alpha Frequency

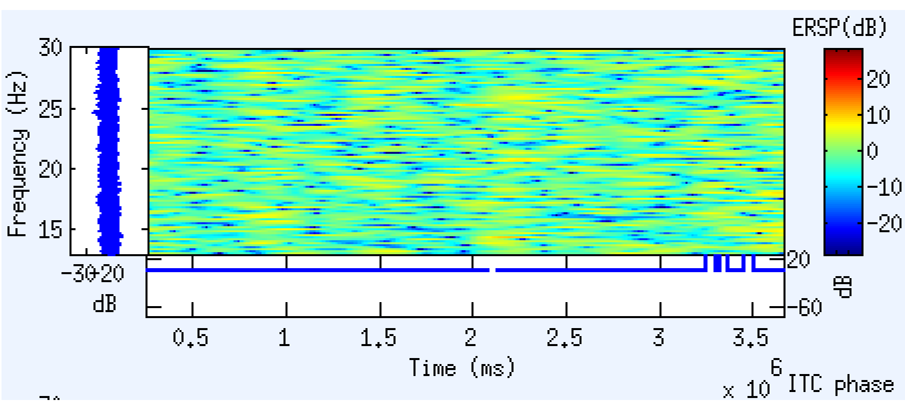


Figure 73: Participant 3 EEG Channel 46 Beta Frequency

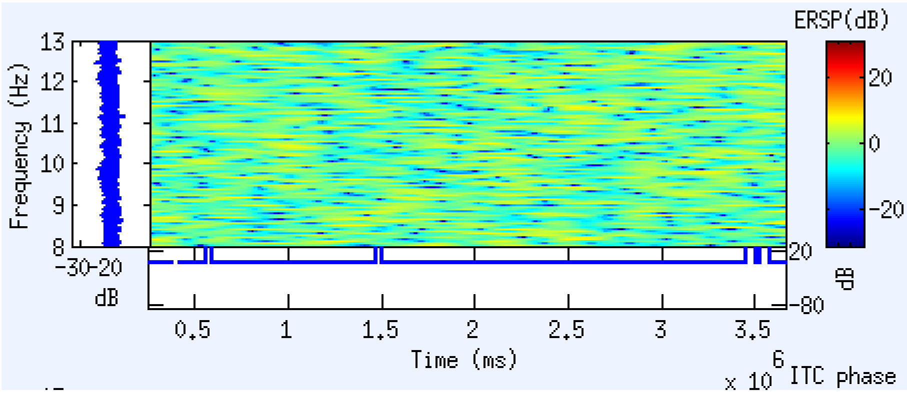


Figure 74: Participant 3 EEG Channel 47 Alpha Frequency

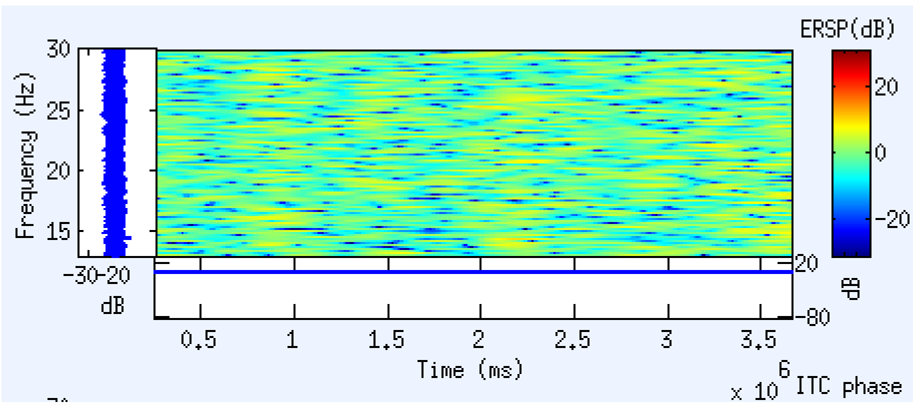


Figure 75: Participant 3 EEG Channel 47 Beta Frequency

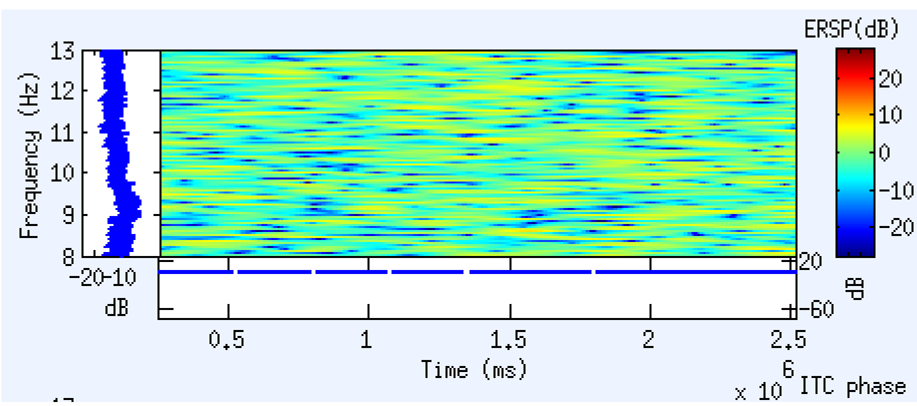


Figure 76: Participant 4 EEG Channel 4 Alpha Frequency

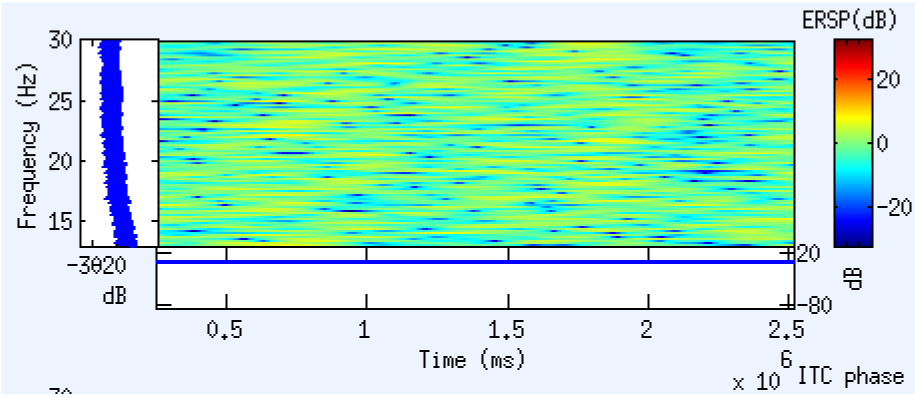


Figure 77: Participant 4 EEG Channel 4 Beta Frequency

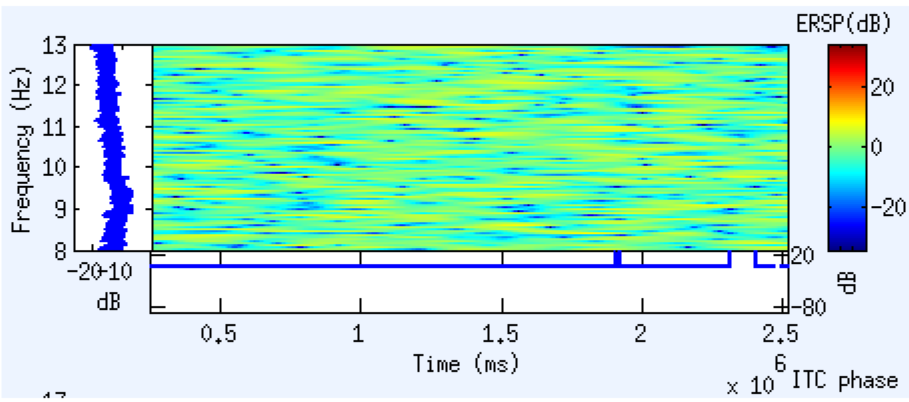


Figure 78: Participant 4 EEG Channel 11 Alpha Frequency

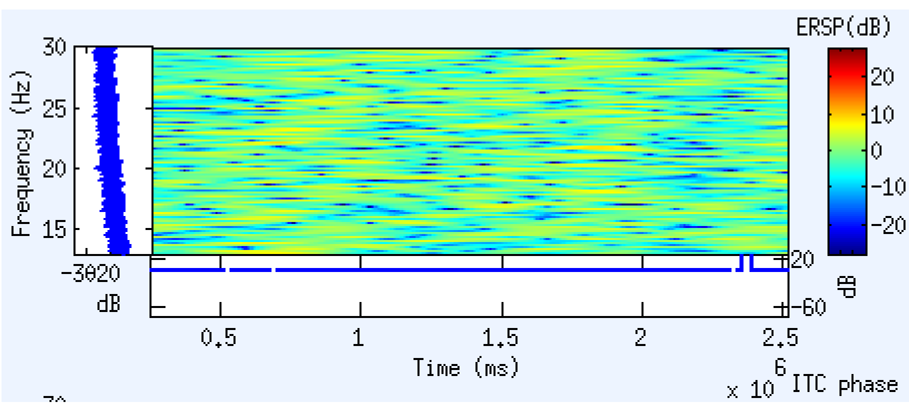


Figure 79: Participant 4 EEG Channel 11 Beta Frequency

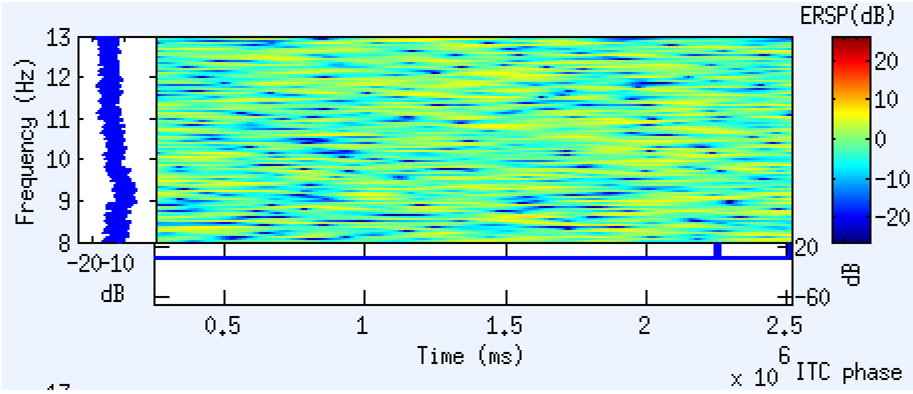


Figure 80: Participant 4 EEG Channel 38 Alpha Frequency

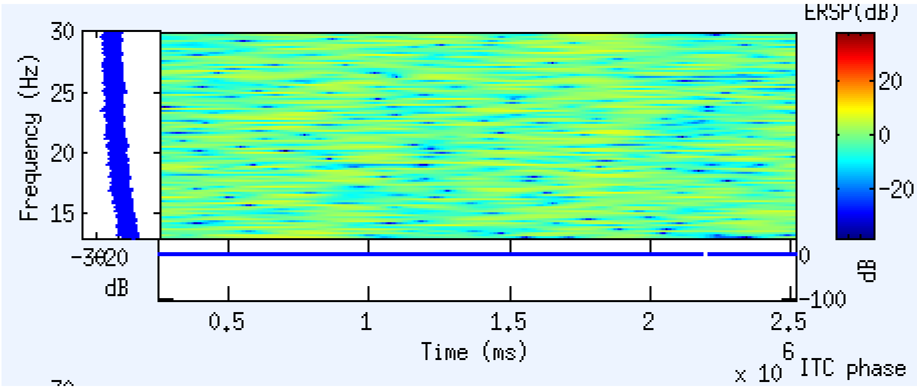


Figure 81: Participant 4 EEG Channel 38 Beta Frequency

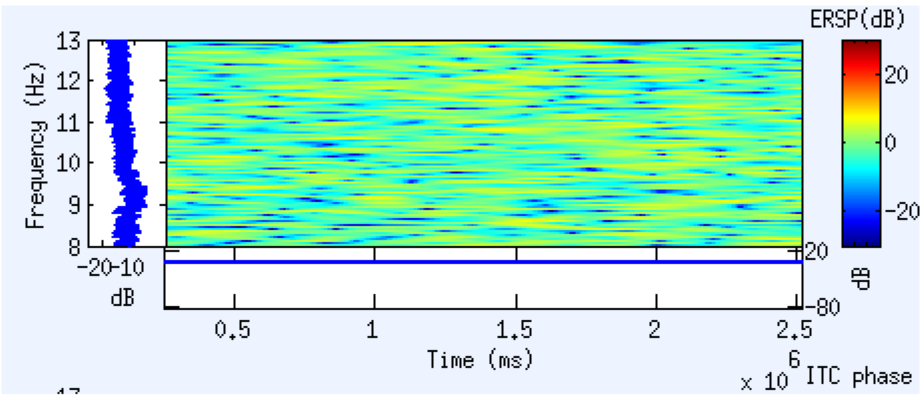


Figure 82: Participant 4 EEG Channel 39 Alpha Frequency

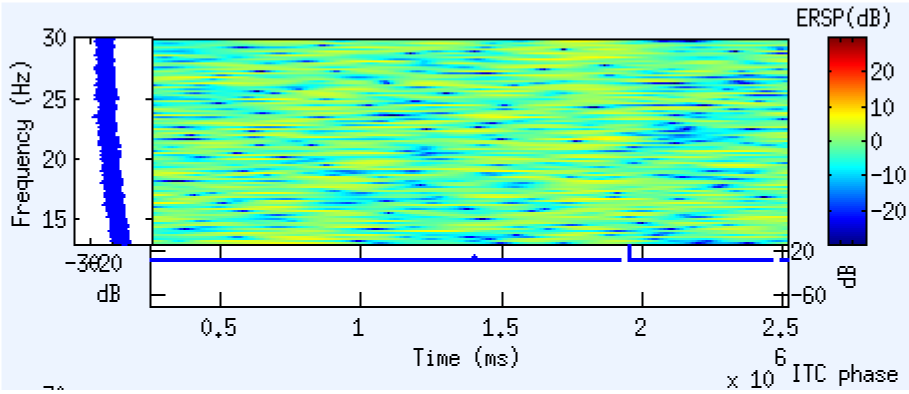


Figure 83: Participant 4 EEG Channel 39 Beta Frequency

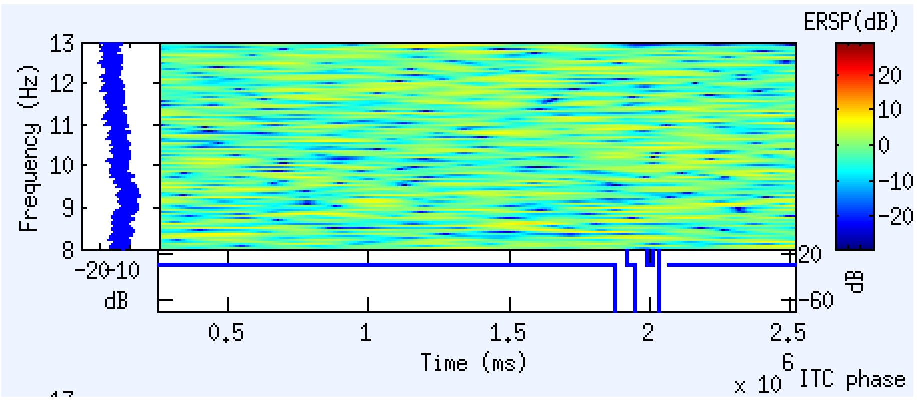


Figure 84: Participant 4 EEG Channel 46 Alpha Frequency

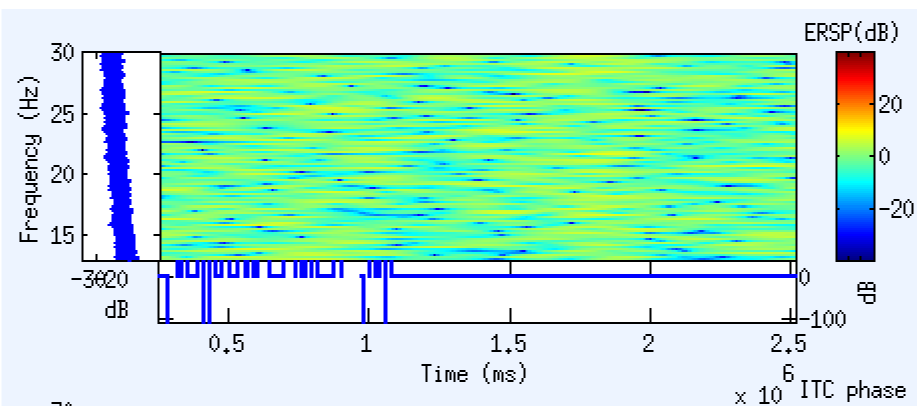


Figure 85: Participant 4 EEG Channel 46 Beta Frequency

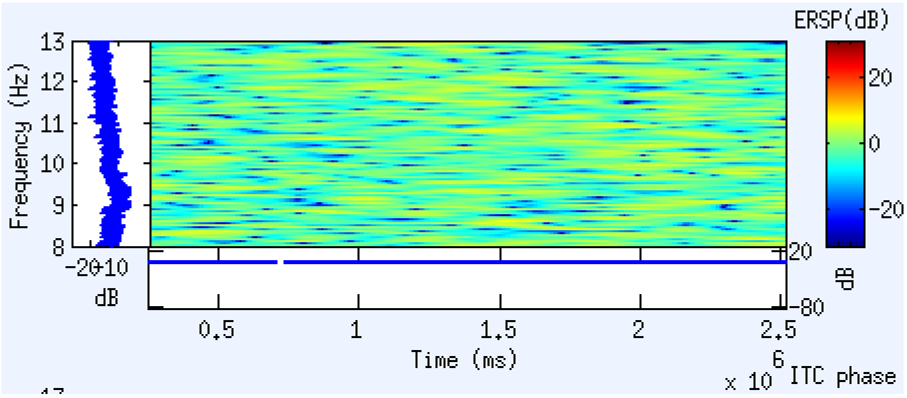


Figure 86: Participant 4 EEG Channel 47 Alpha Frequency

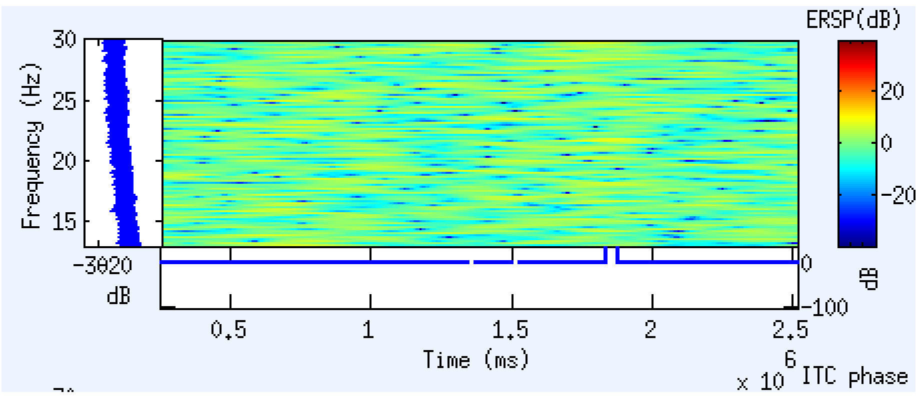


Figure 87: Participant 4 EEG Channel 47 Beta Frequency

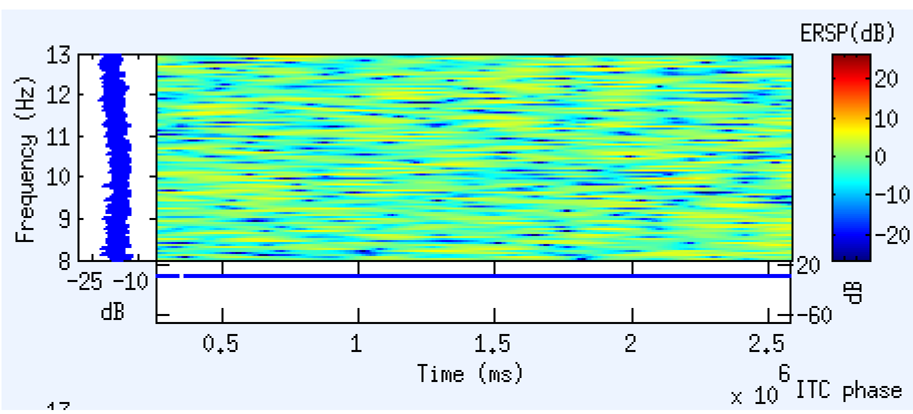


Figure 88: Participant 5 EEG Channel 4 Alpha Frequency

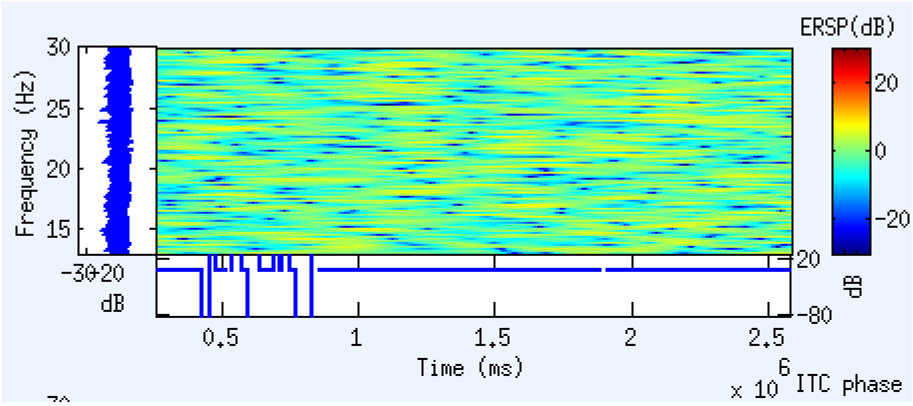


Figure 89: Participant 5 EEG Channel 4 Beta Frequency

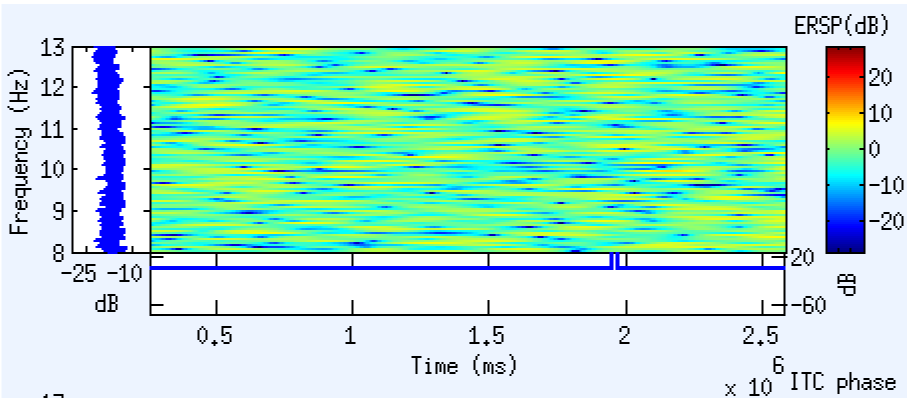


Figure 90: Participant 5 EEG Channel 11 Alpha Frequency

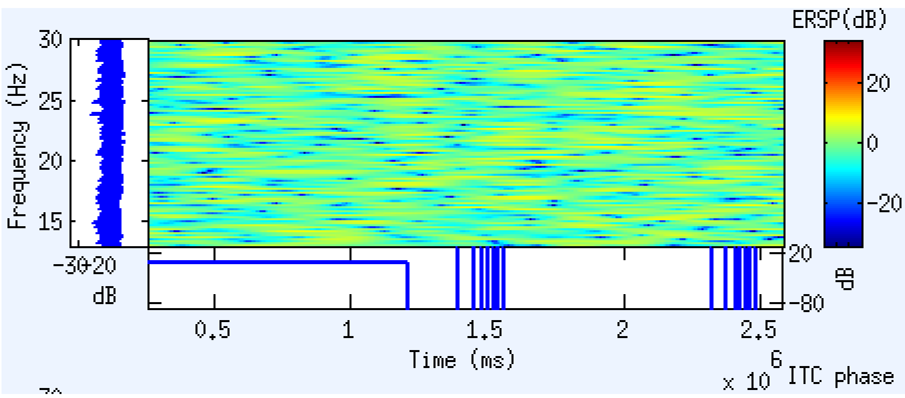


Figure 91: Participant 5 EEG Channel 11 Beta Frequency

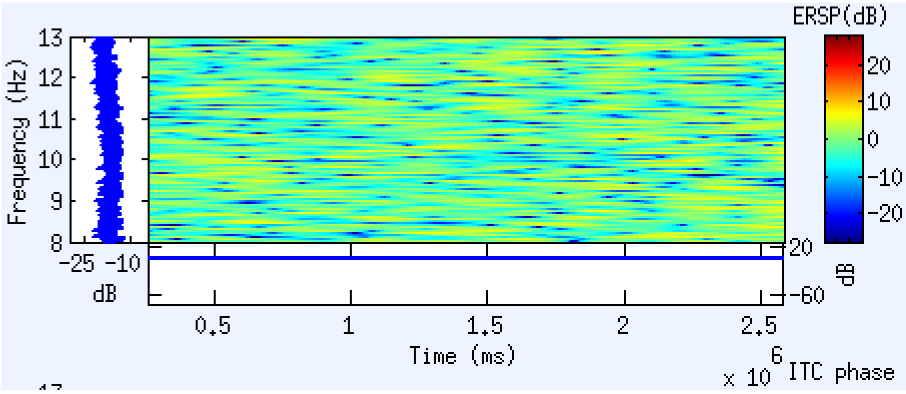


Figure 92: Participant 5 EEG Channel 38 Alpha Frequency

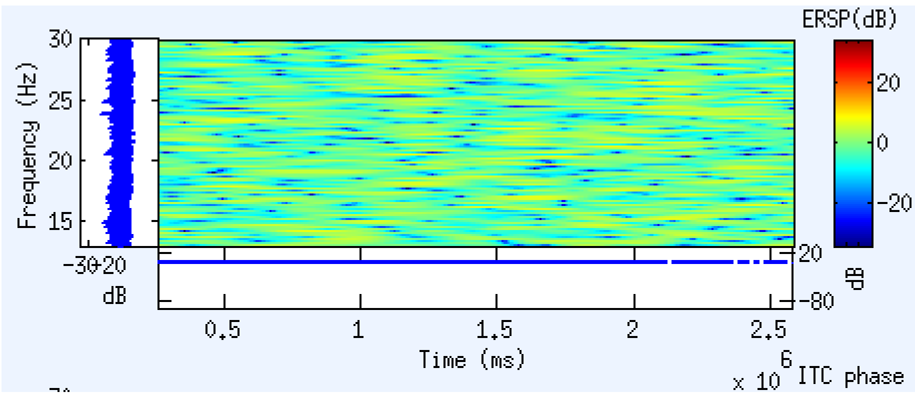


Figure 93: Participant 5 EEG Channel 38 Beta Frequency

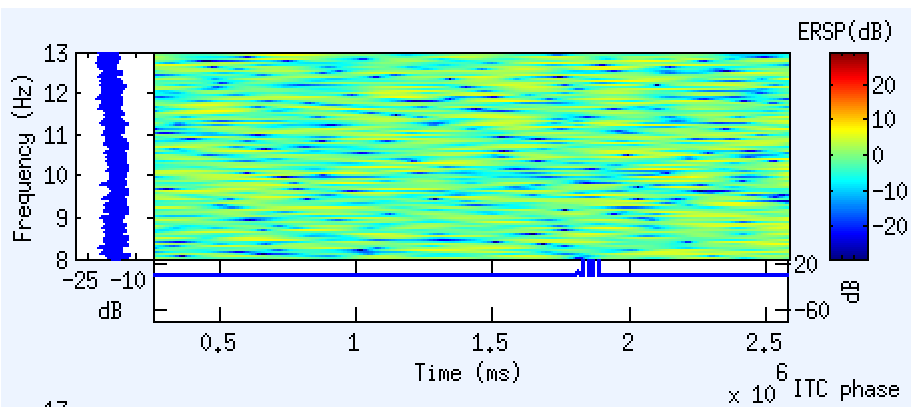


Figure 94: Participant 5 EEG Channel 39 Alpha Frequency

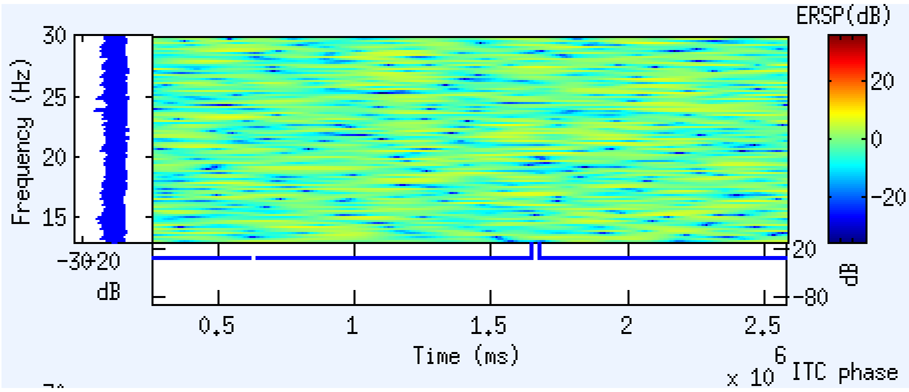


Figure 95: Participant 5 EEG Channel 39 Beta Frequency

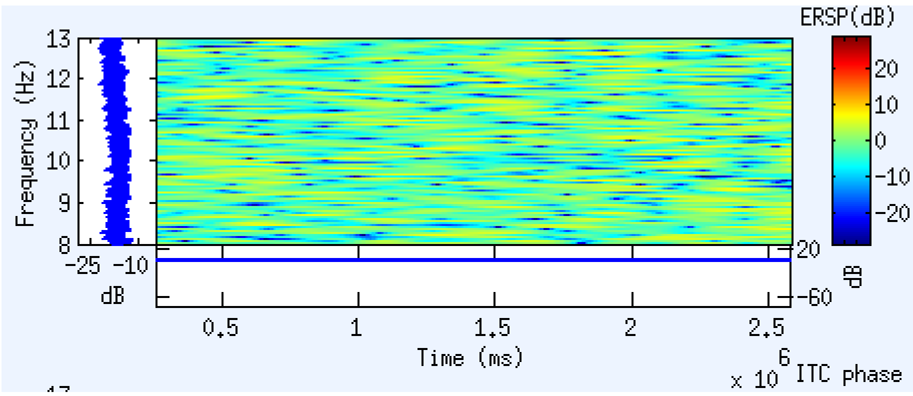


Figure 96: Participant 5 EEG Channel 46 Alpha Frequency

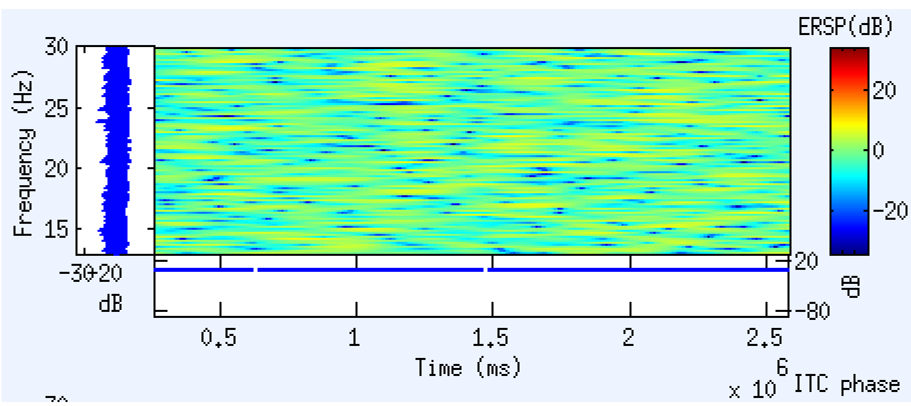


Figure 97: Participant 5 EEG Channel 46 Beta Frequency

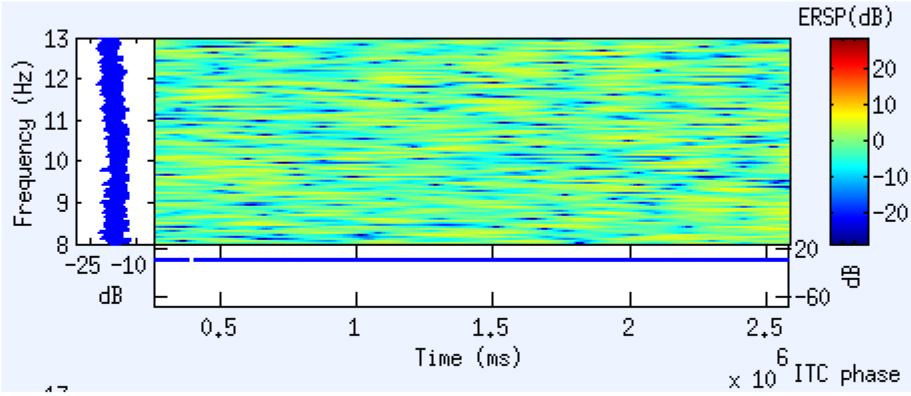


Figure 98: Participant 5 EEG Channel 47 Alpha Frequency

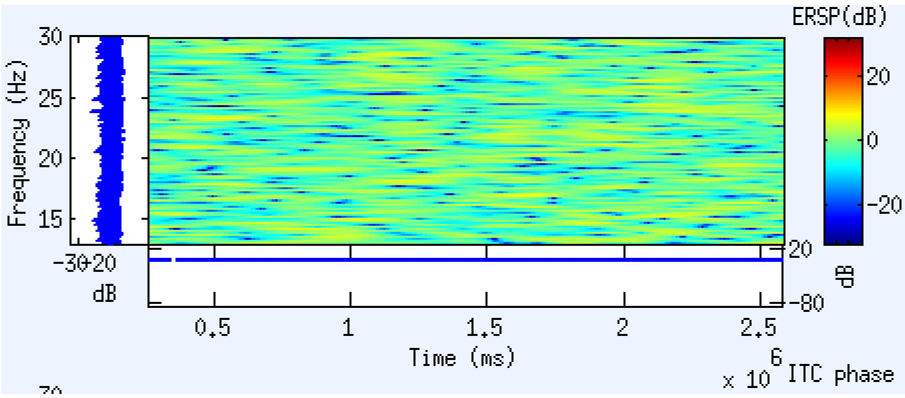


Figure 99: Participant 5 EEG Channel 47 Beta Frequency

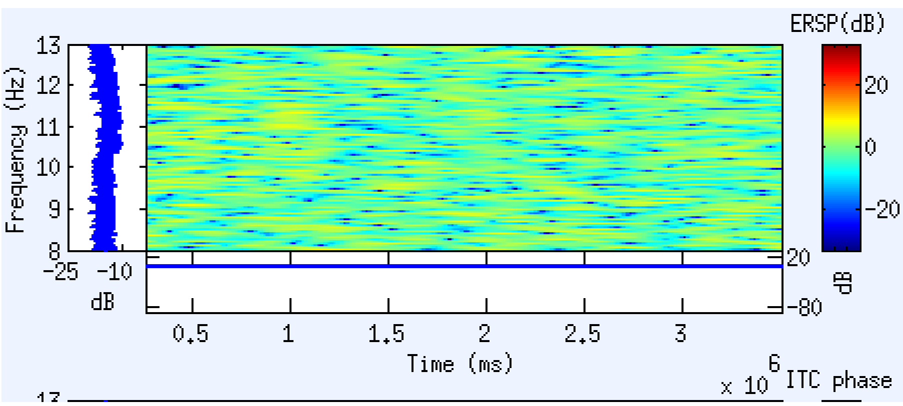


Figure 100: Participant 6 EEG Channel 4 Alpha Frequency

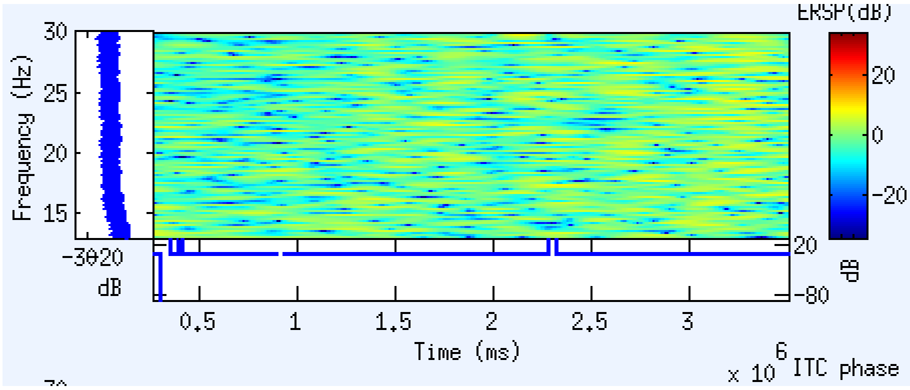


Figure 101: Participant 6 EEG Channel 4 Beta Frequency

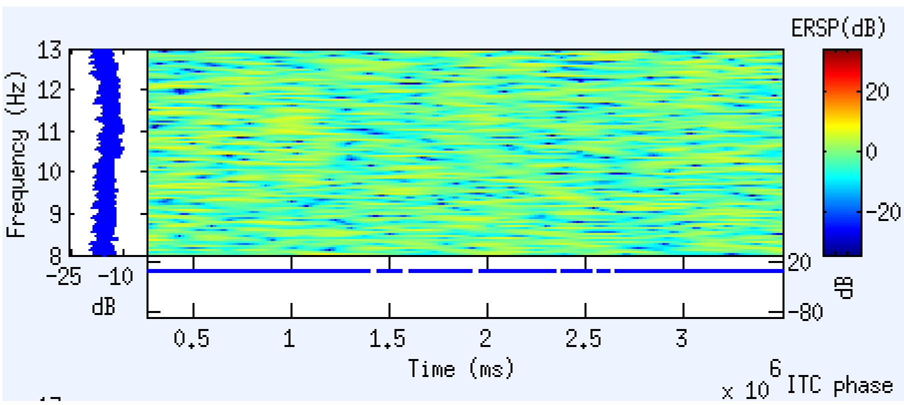


Figure 102: Participant 6 EEG Channel 11 Alpha Frequency

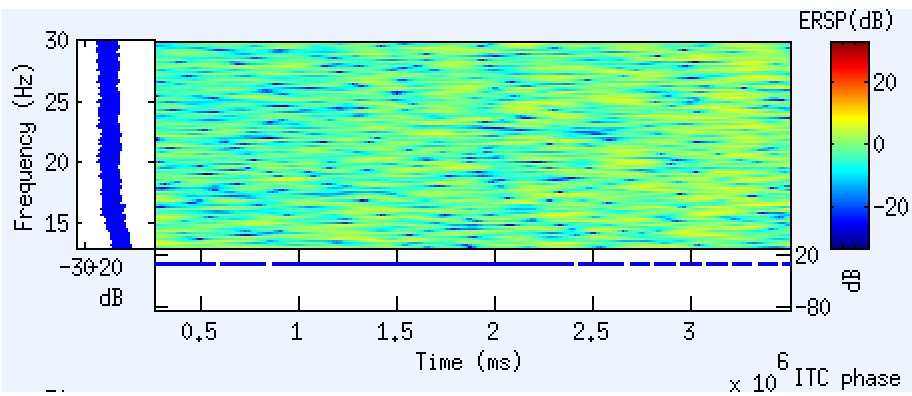


Figure 103: Participant 6 EEG Channel 11 Beta Frequency

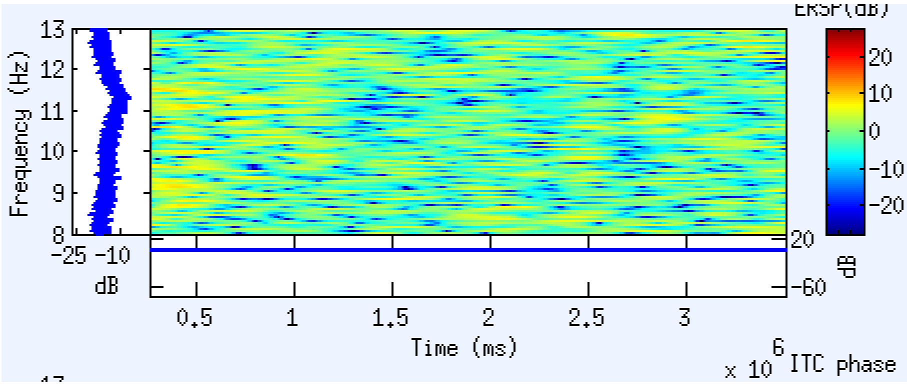


Figure 104: Participant 6 EEG Channel 38 Alpha Frequency

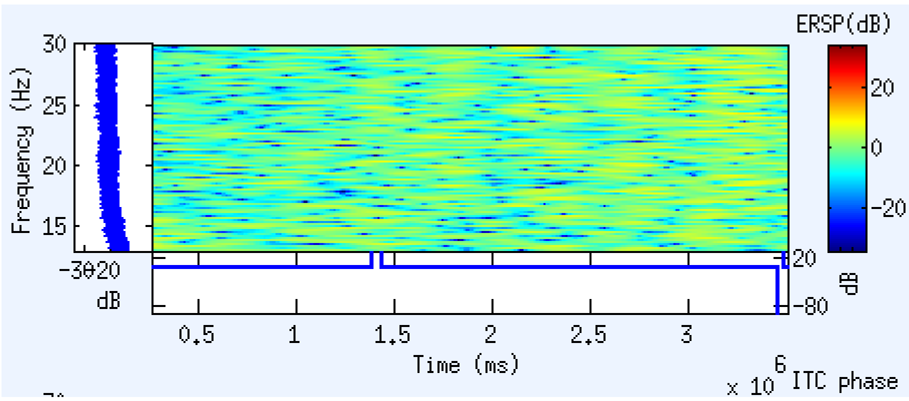


Figure 105: Participant 6 EEG Channel 38 Beta Frequency

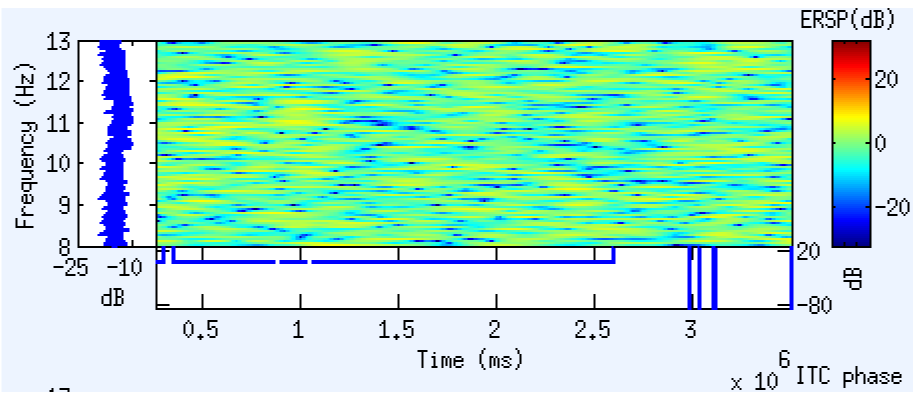


Figure 106: Participant 6 EEG Channel 39 Alpha Frequency

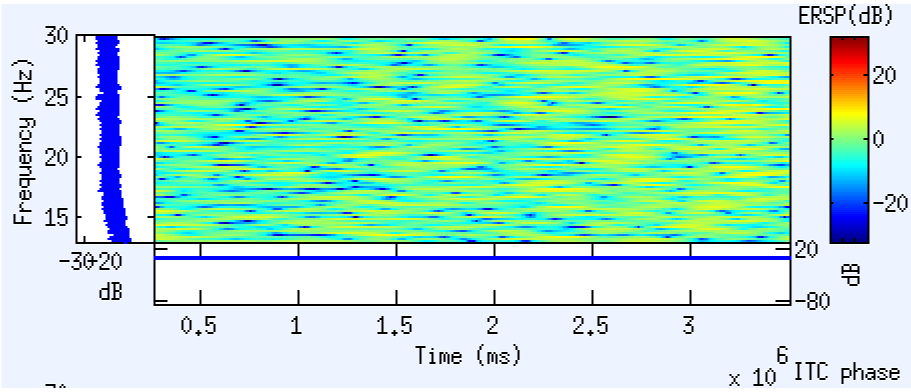


Figure 107: Participant 6 EEG Channel 39 Beta Frequency

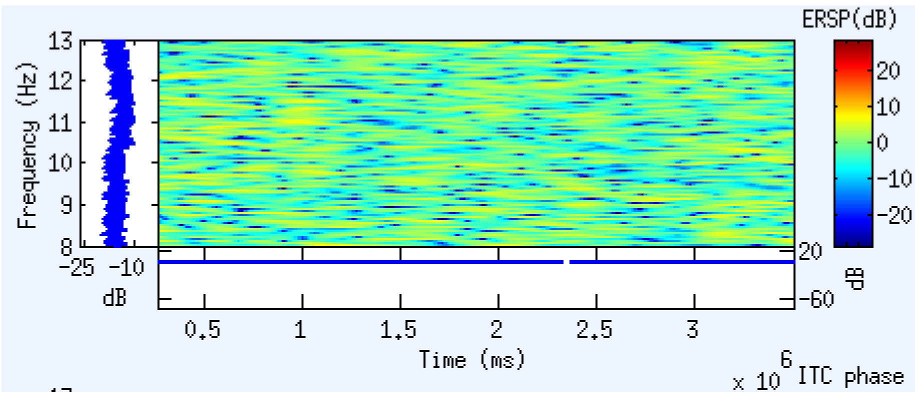


Figure 108: Participant 6 EEG Channel 46 Alpha Frequency

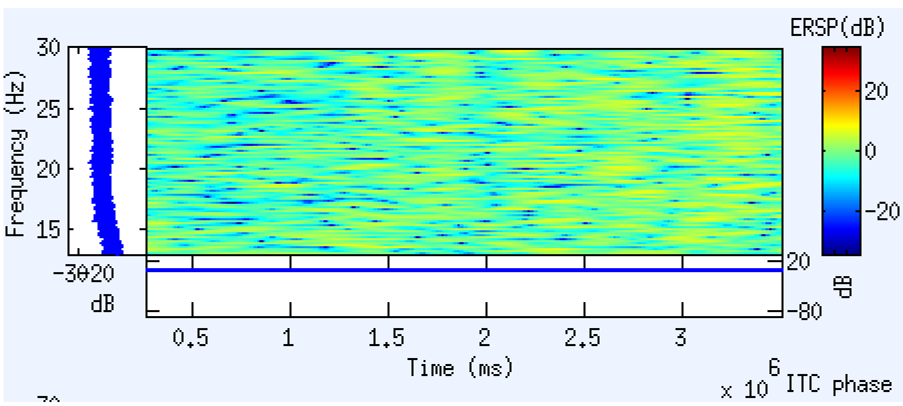


Figure 109: Participant 6 EEG Channel 46 Beta Frequency

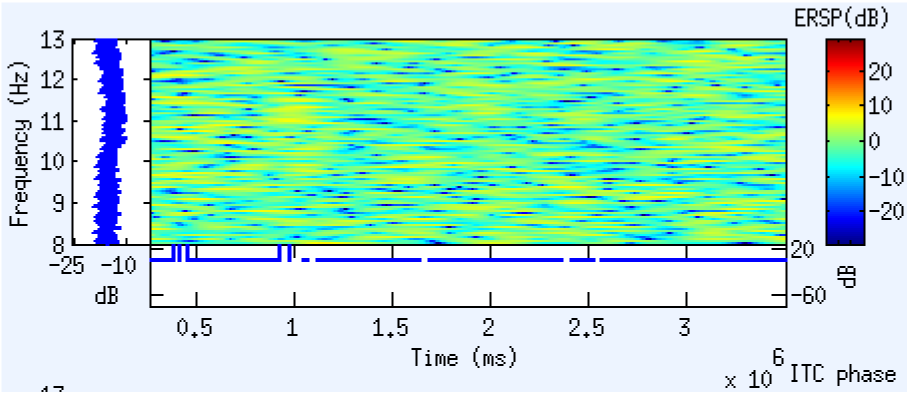


Figure 110: Participant 6 EEG Channel 47 Alpha Frequency

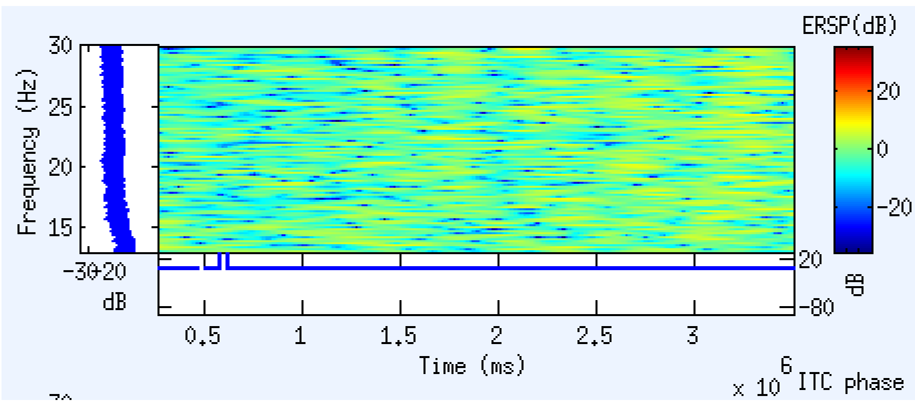


Figure 111: Participant 6 EEG Channel 47 Beta Frequency

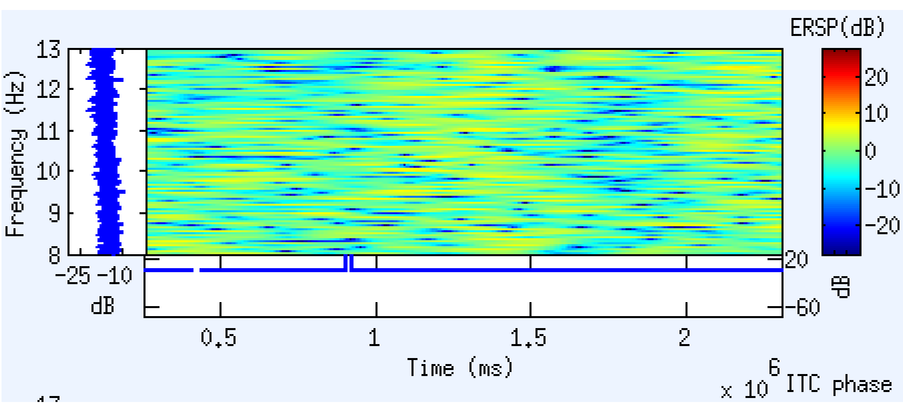


Figure 112: Participant 7 EEG Channel 4 Alpha Frequency

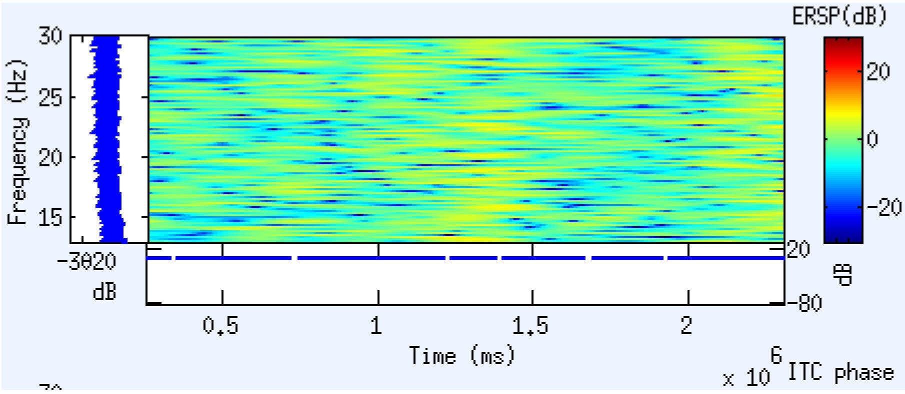


Figure 113: Participant 7 EEG Channel 4 Beta Frequency

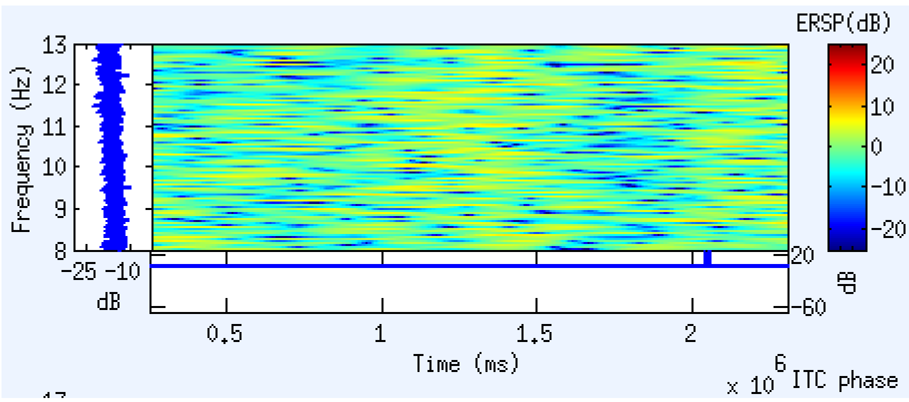


Figure 114: Participant 7 EEG Channel 11 Alpha Frequency

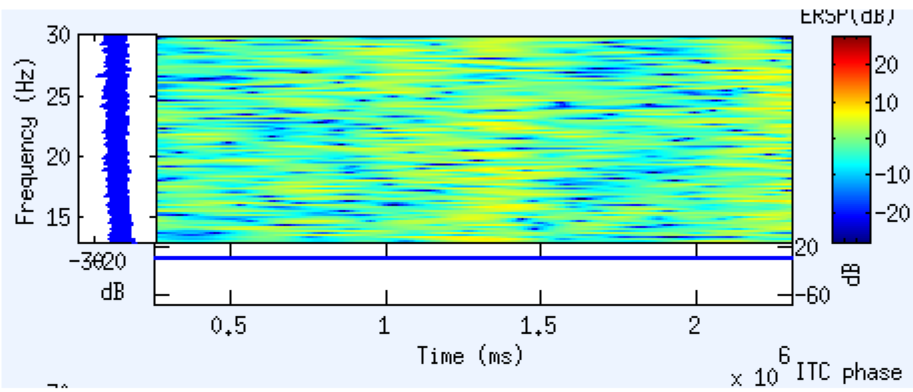


Figure 115: Participant 7 EEG Channel 11 Beta Frequency

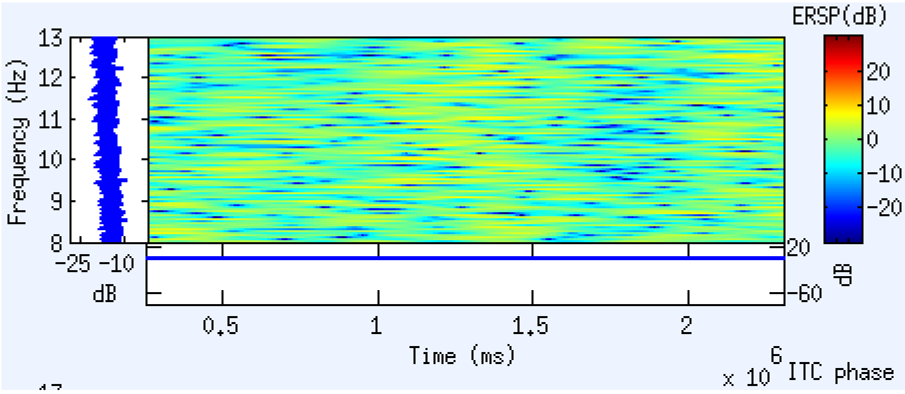


Figure 116: Participant 7 EEG Channel 38 Alpha Frequency

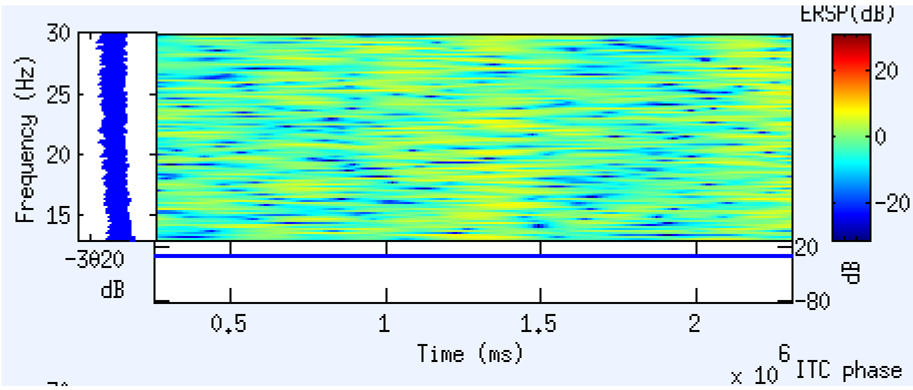


Figure 117: Participant 7 EEG Channel 38 Beta Frequency

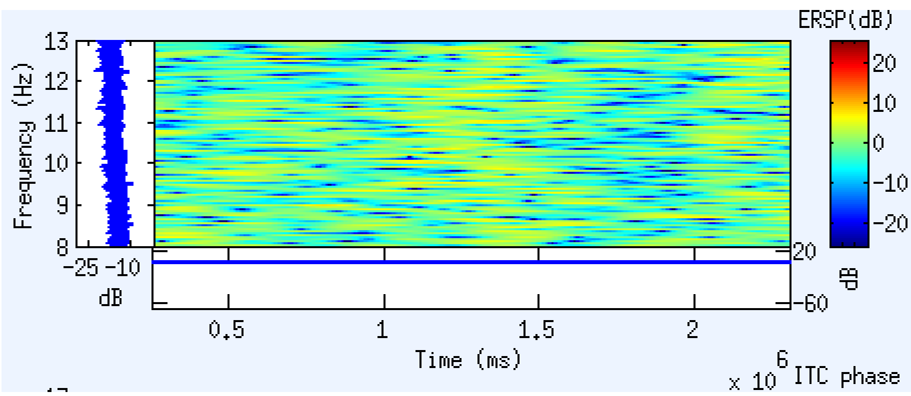


Figure 118: Participant 7 EEG Channel 39 Alpha Frequency

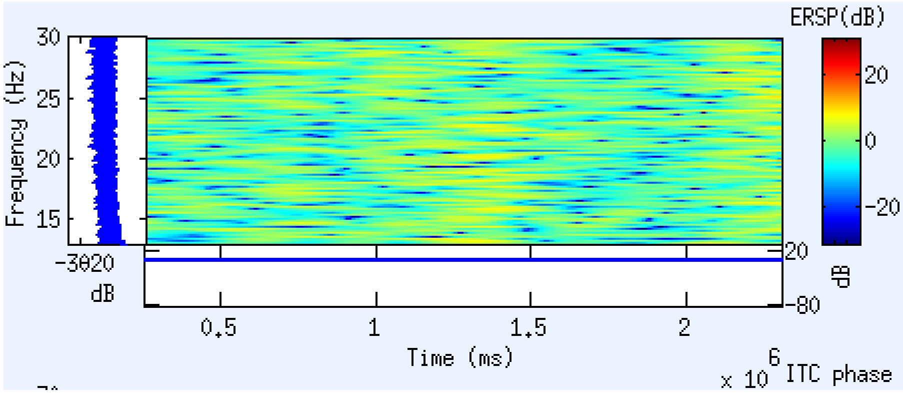


Figure 119: Participant 7 EEG Channel 39 Beta Frequency

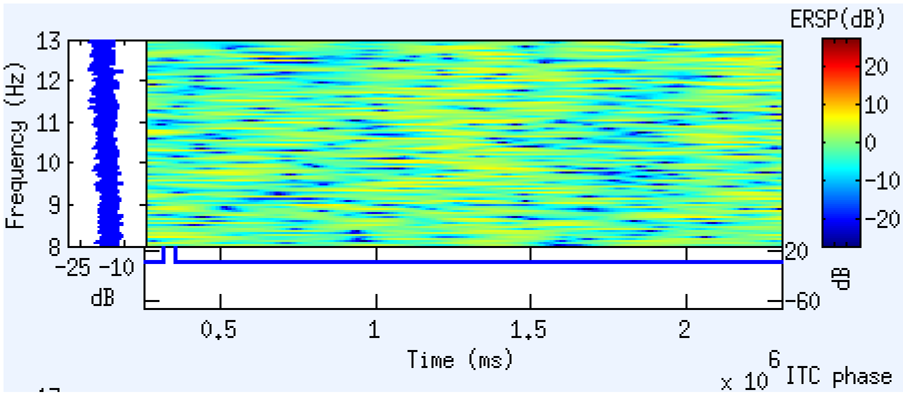


Figure 120: Participant 7 EEG Channel 46 Alpha Frequency

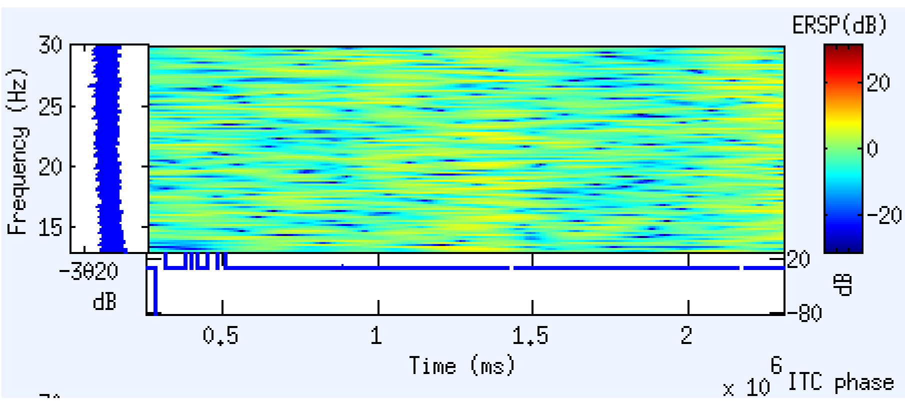


Figure 121: Participant 7 EEG Channel 46 Beta Frequency

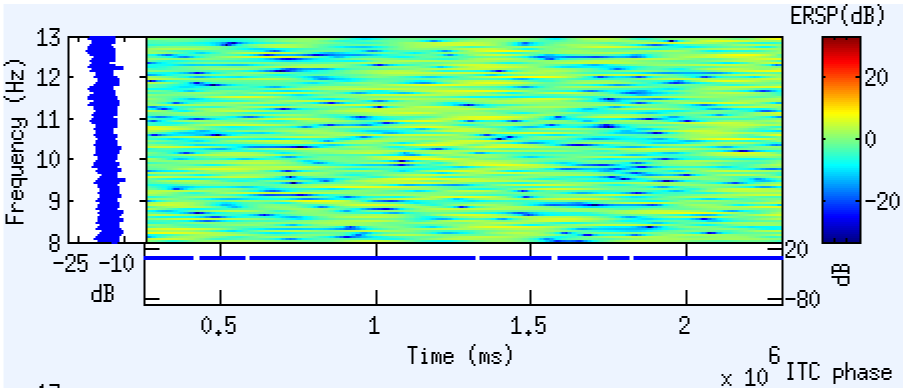


Figure 122: Participant 7 EEG Channel 47 Alpha Frequency

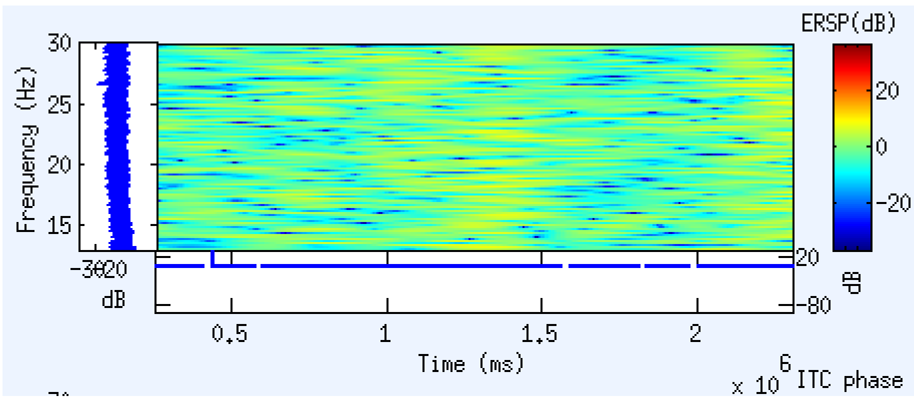


Figure 123: Participant 7 EEG Channel 47 Beta Frequency

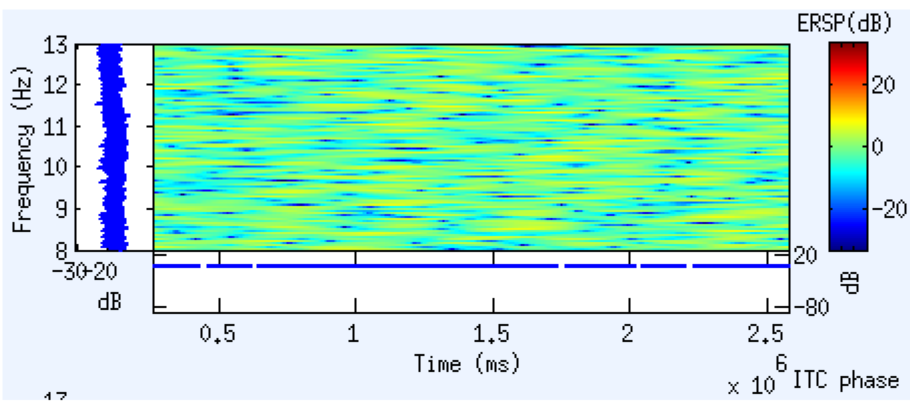


Figure 124: Participant 8 EEG Channel 4 Alpha Frequency

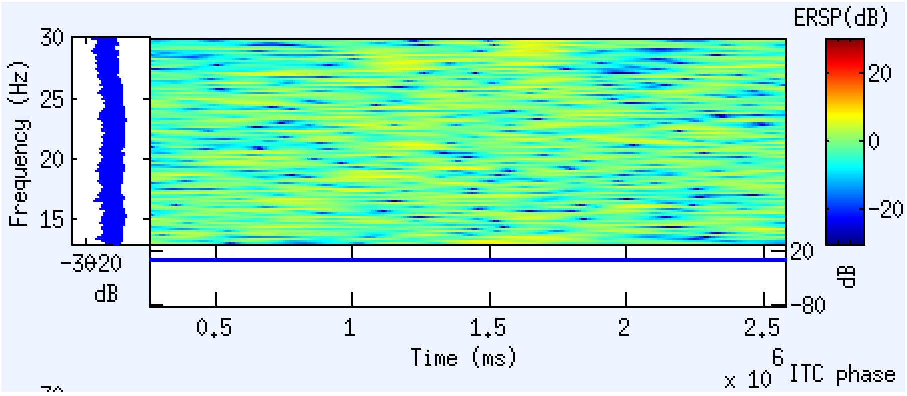


Figure 125: Participant 8 EEG Channel 4 Beta Frequency

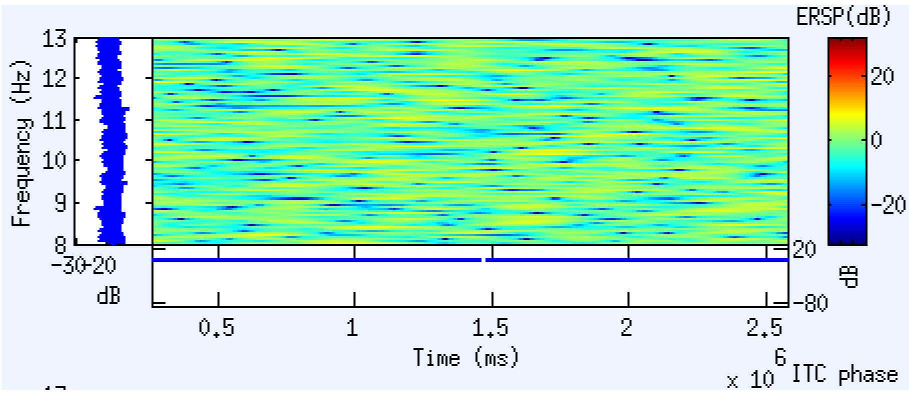


Figure 126: Participant 8 EEG Channel 11 Alpha Frequency

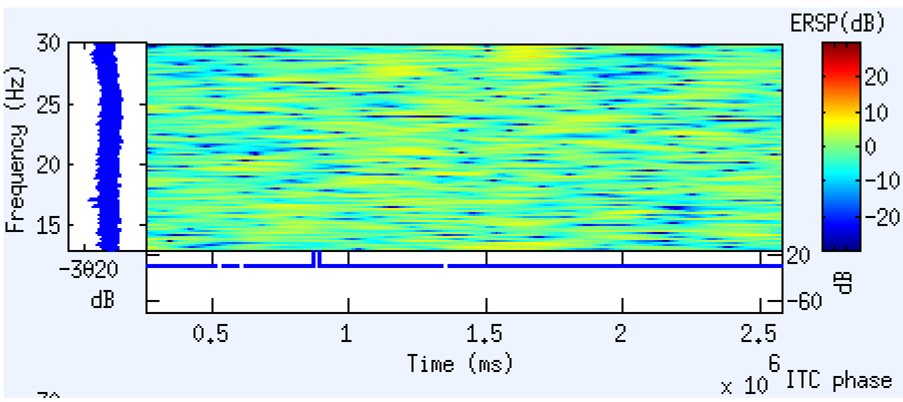


Figure 127: Participant 8 EEG Channel 11 Beta Frequency

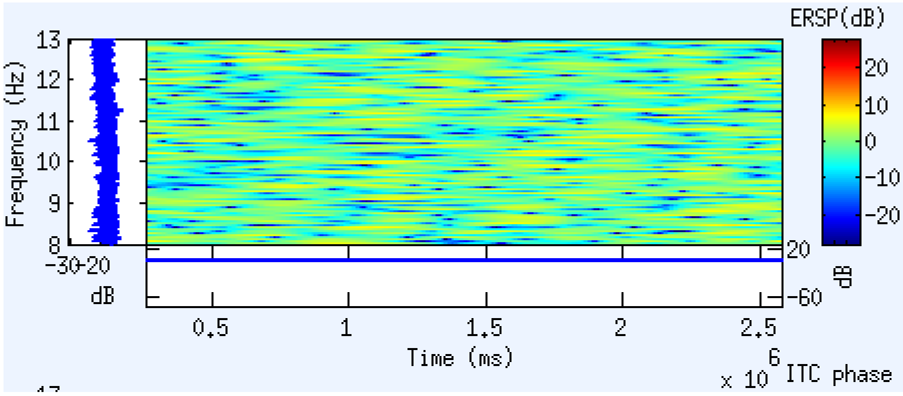


Figure 128: Participant 8 EEG Channel 38 Alpha Frequency

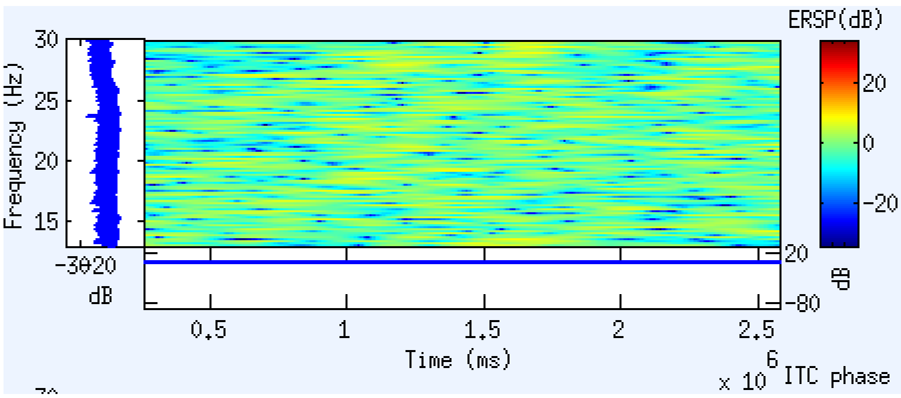


Figure 129: Participant 8 EEG Channel 38 Beta Frequency

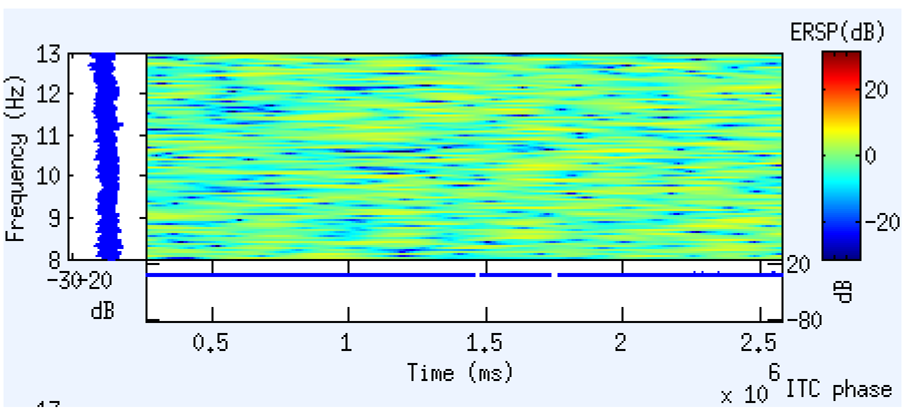


Figure 130: Participant 8 EEG Channel 39 Alpha Frequency

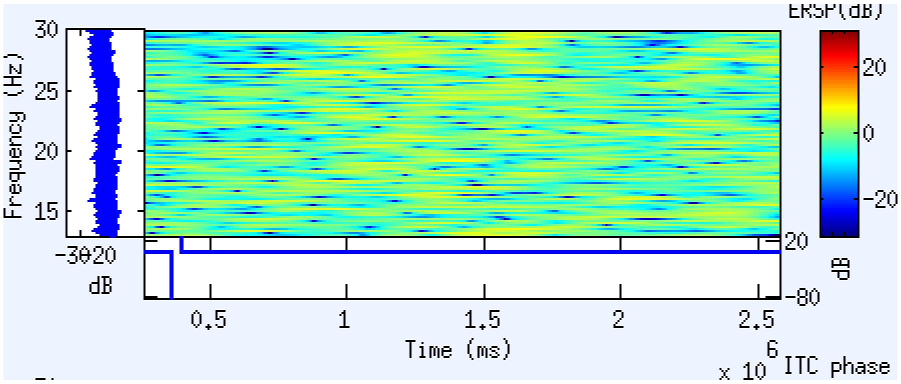


Figure 131: Participant 8 EEG Channel 39 Beta Frequency

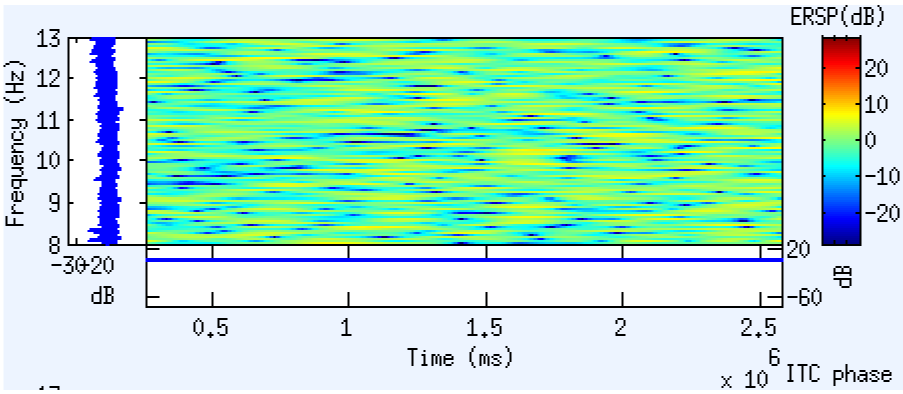


Figure 132: Participant 8 EEG Channel 46 Alpha Frequency

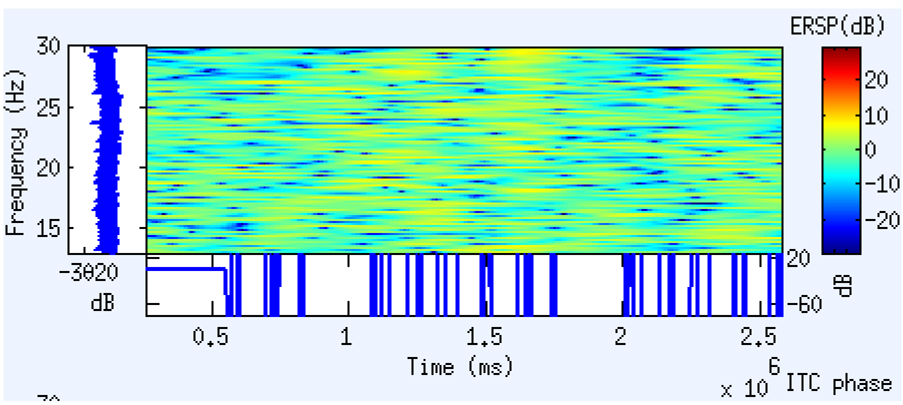


Figure 133: Participant 8 EEG Channel 46 Beta Frequency

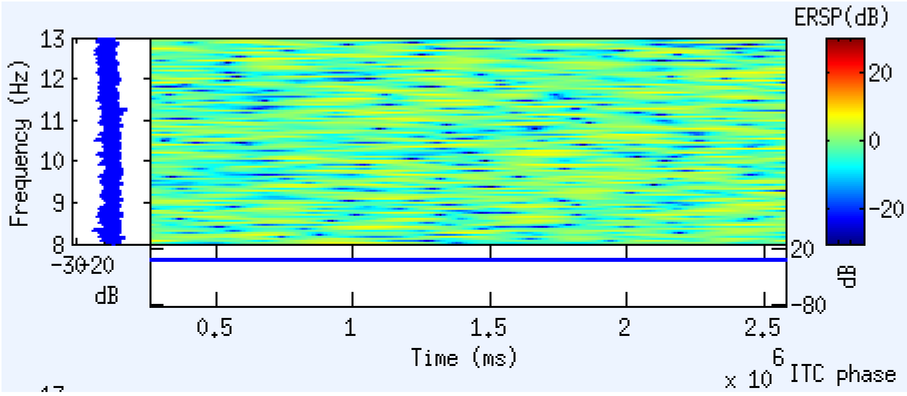


Figure 134: Participant 8 EEG Channel 47 Alpha Frequency

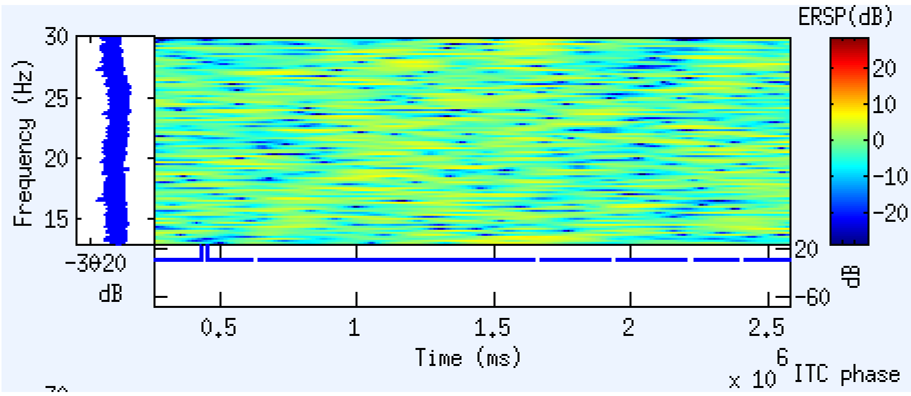


Figure 135: Participant 8 EEG Channel 47 Beta Frequency

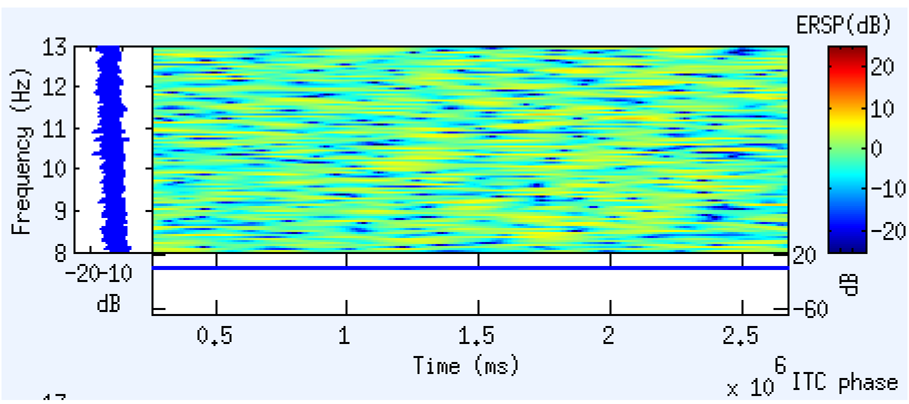


Figure 136: Participant 9 EEG Channel 4 Alpha Frequency

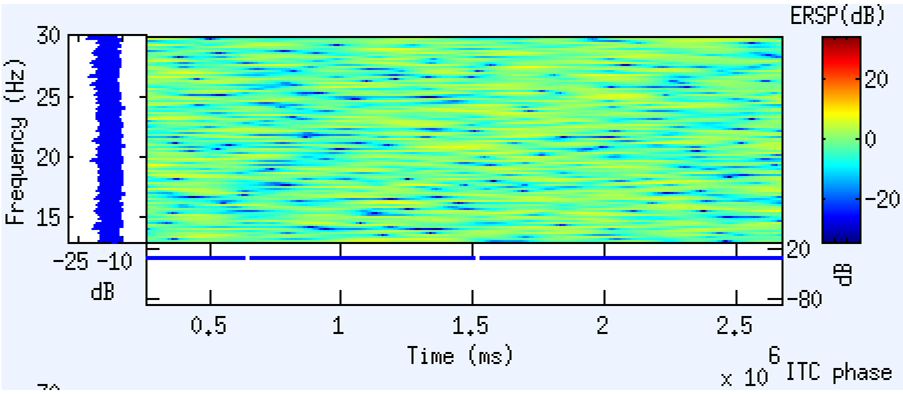


Figure 137: Participant 9 EEG Channel 4 Beta Frequency

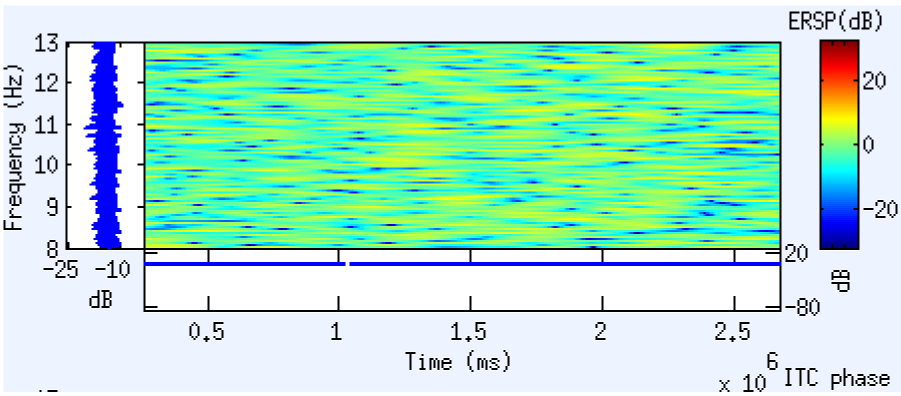


Figure 138: Participant 9 EEG Channel 11 Alpha Frequency

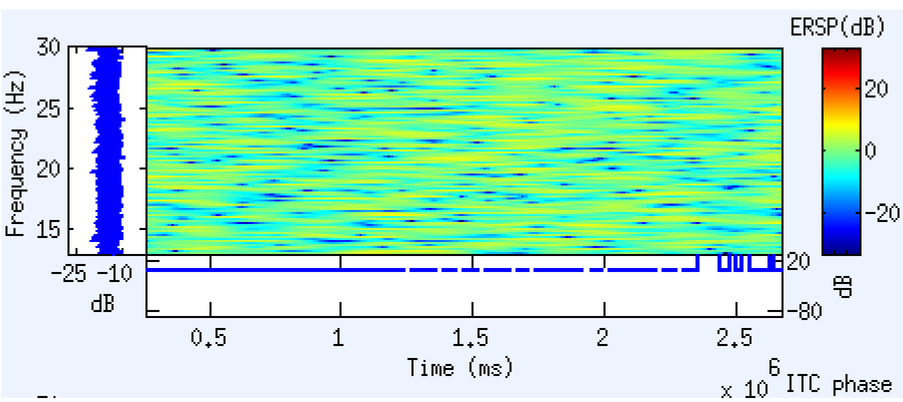


Figure 139: Participant 9 EEG Channel 11 Beta Frequency

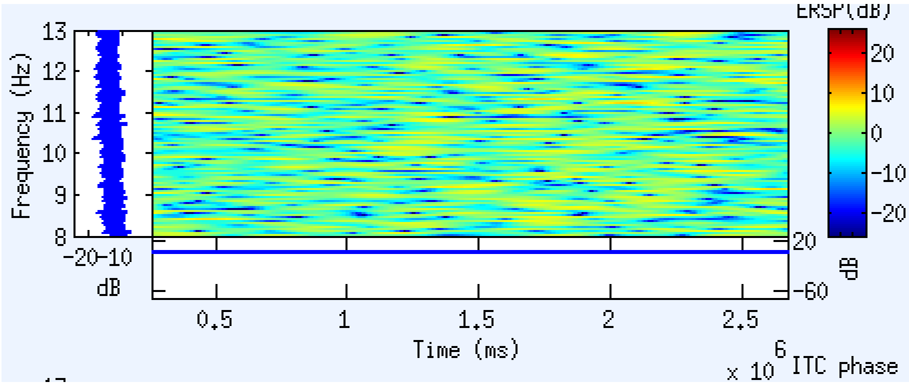


Figure 140: Participant 9 EEG Channel 38 Alpha Frequency

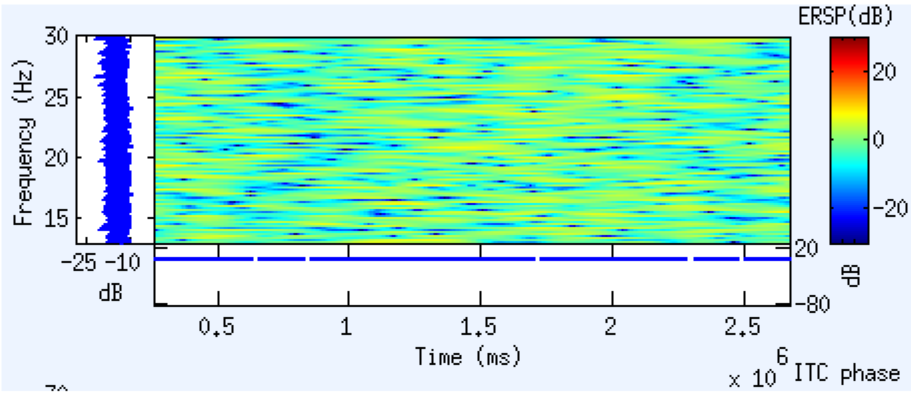


Figure 141: Participant 9 EEG Channel 38 Beta Frequency

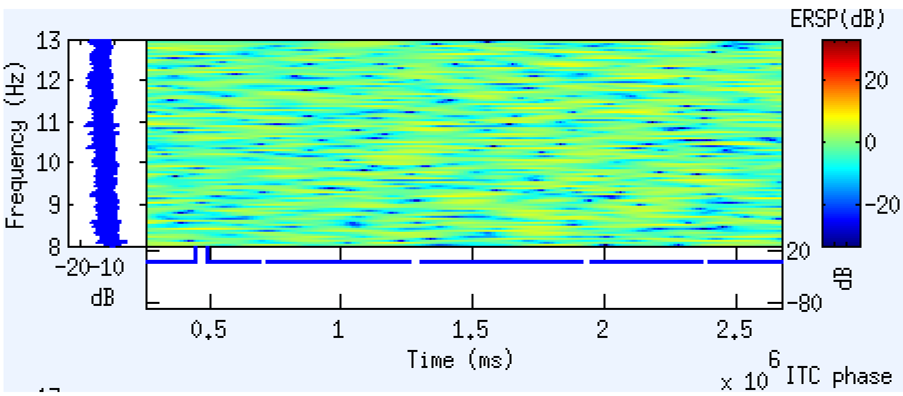


Figure 142: Participant 9 EEG Channel 39 Alpha Frequency

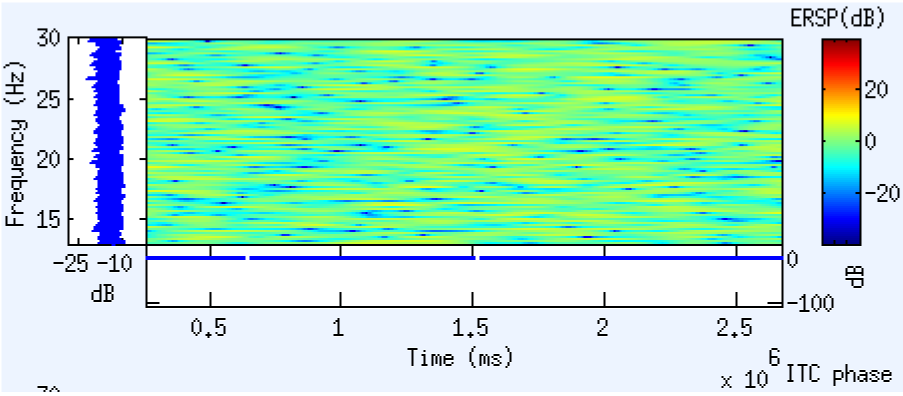


Figure 143: Participant 9 EEG Channel 39 Beta Frequency

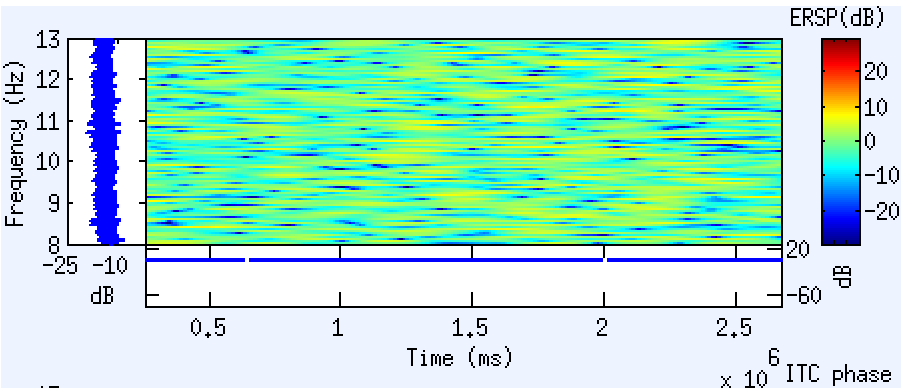


Figure 144: Participant 9 EEG Channel 46 Alpha Frequency

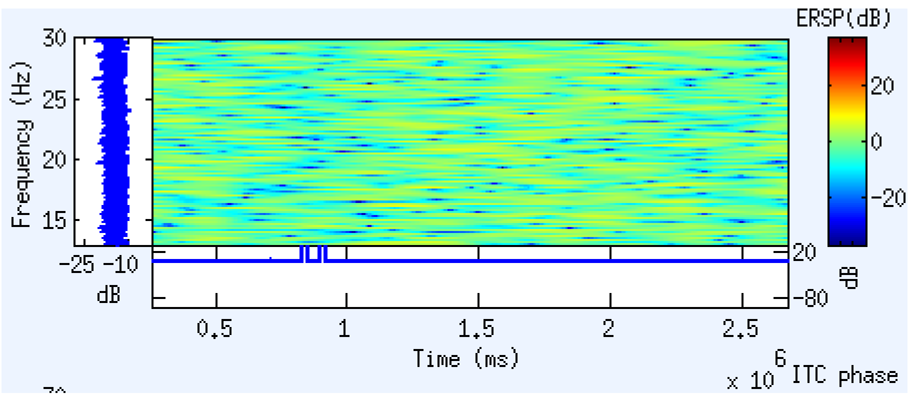


Figure 145: Participant 9 EEG Channel 46 Beta Frequency

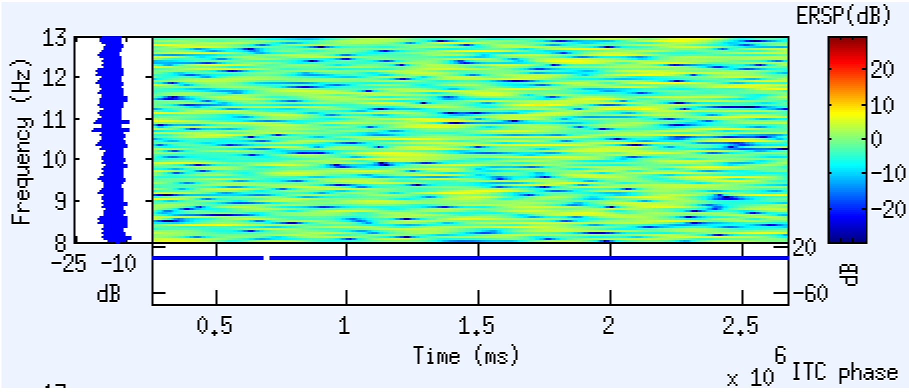


Figure 146: Participant 9 EEG Channel 47 Alpha Frequency

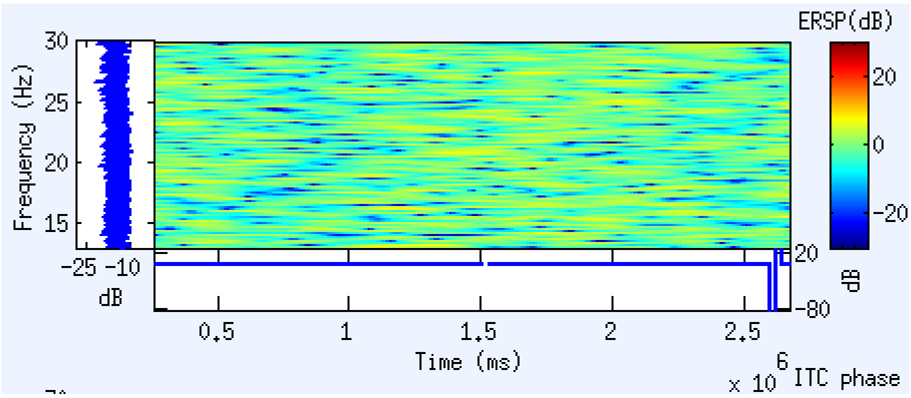


Figure 147: Participant 9 EEG Channel 47 Beta Frequency

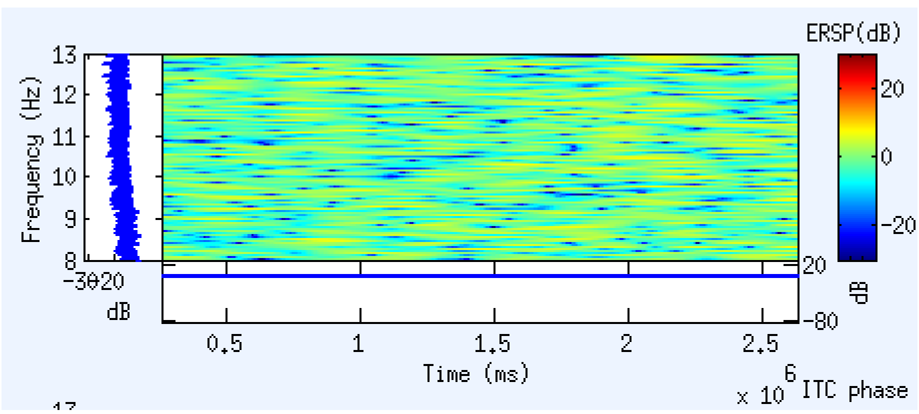


Figure 148: Participant 10 EEG Channel 4 Alpha Frequency

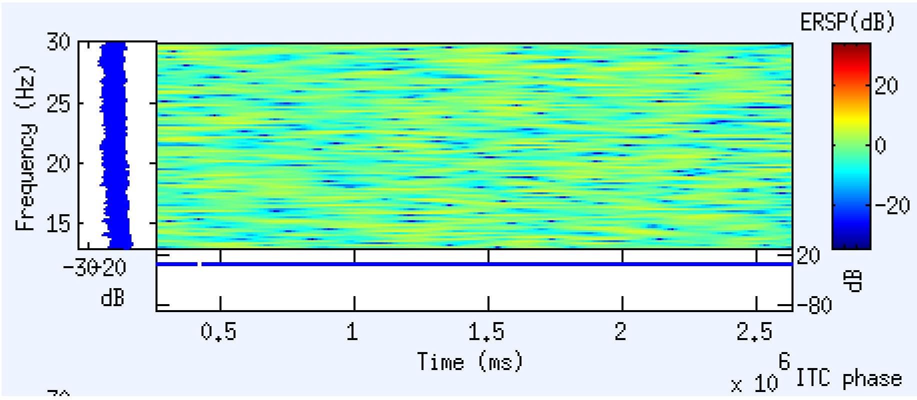


Figure 149: Participant 10 EEG Channel 4 Beta Frequency

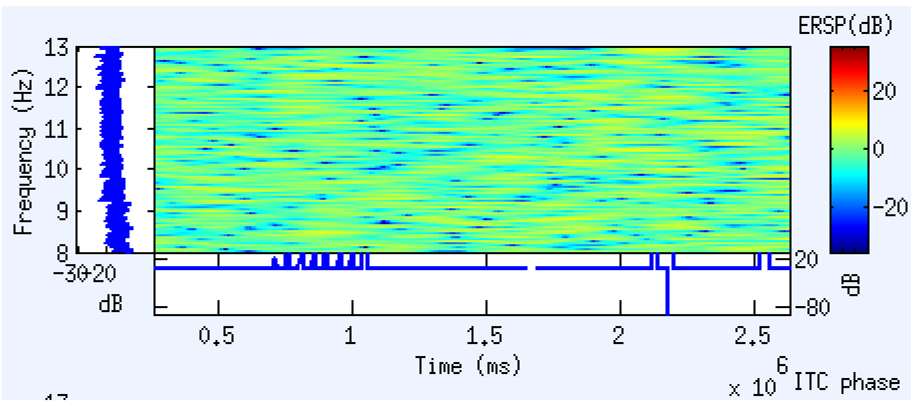


Figure 150: Participant 10 EEG Channel 11 Alpha Frequency

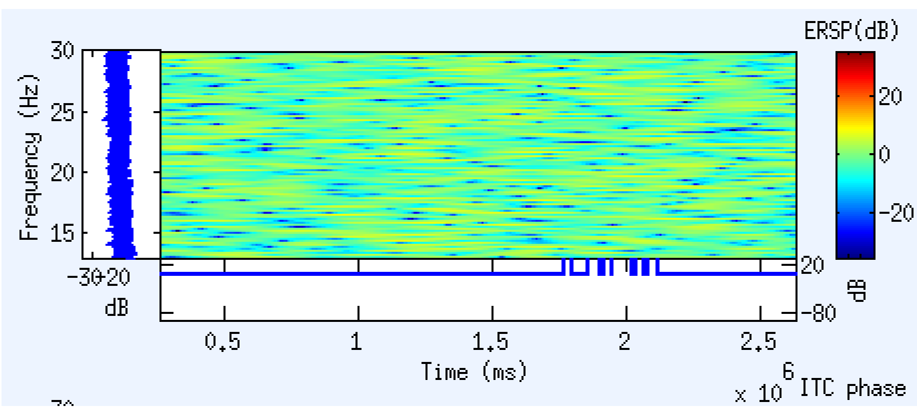


Figure 151: Participant 10 EEG Channel 11 Beta Frequency

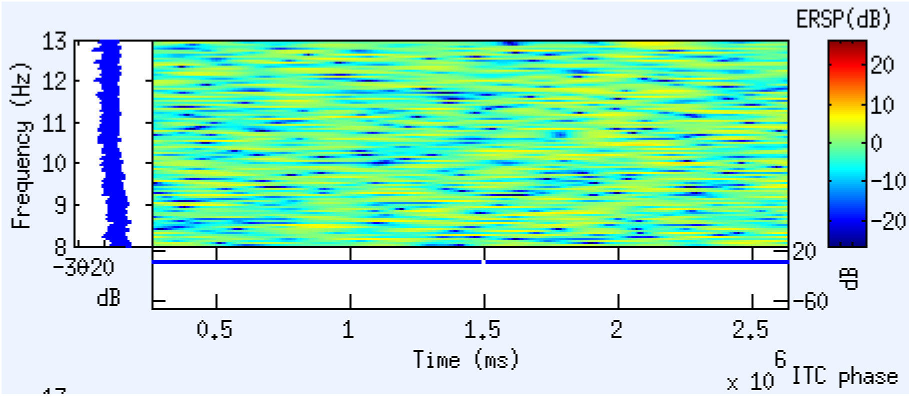


Figure 152: Participant 10 EEG Channel 38 Alpha Frequency

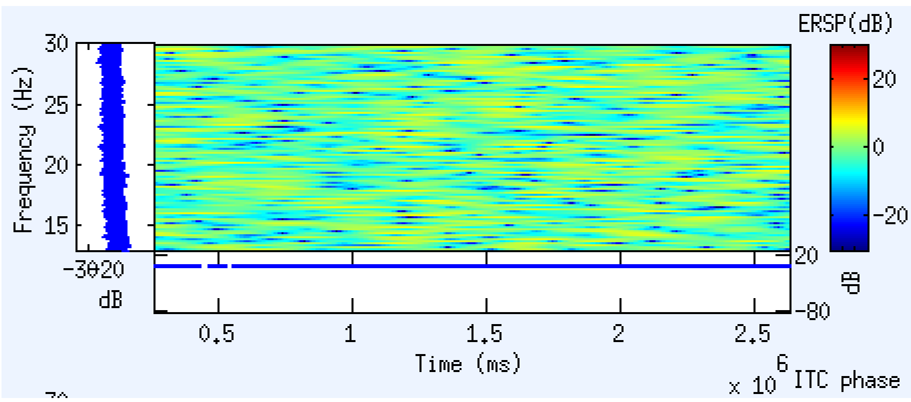


Figure 153: Participant 10 EEG Channel 38 Beta Frequency

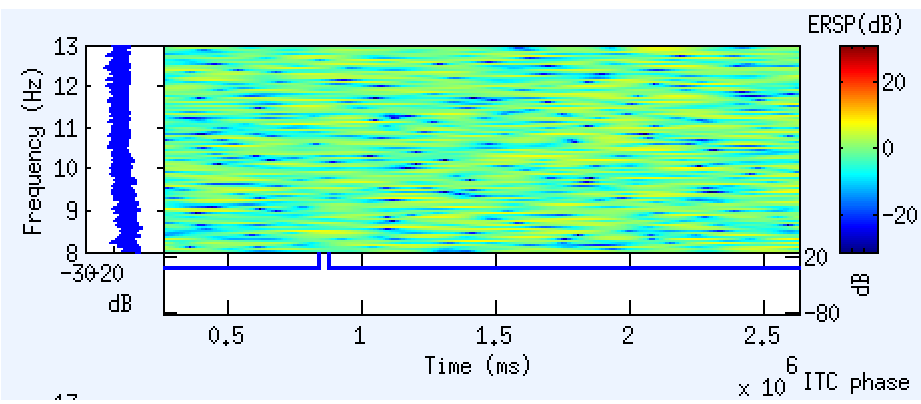


Figure 154: Participant 10 EEG Channel 39 Alpha Frequency

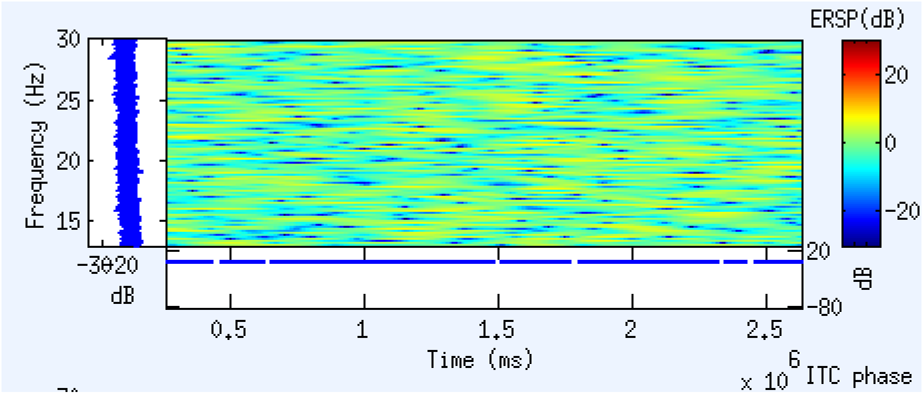


Figure 155: Participant 10 EEG Channel 39 Beta Frequency

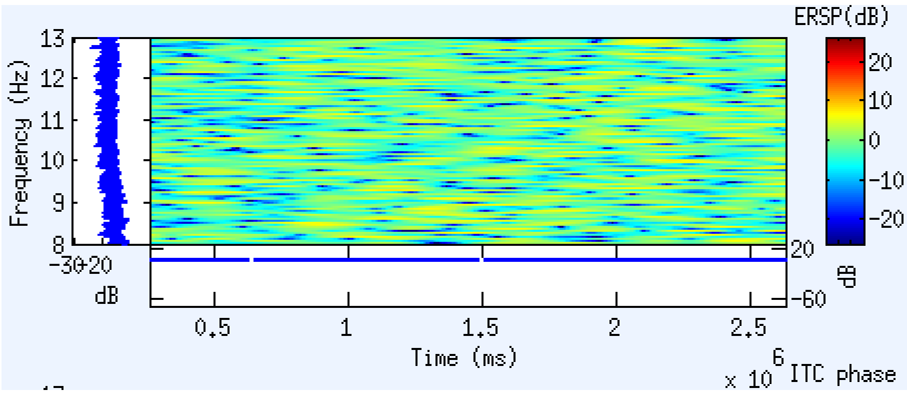


Figure 156: Participant 10 EEG Channel 46 Alpha Frequency

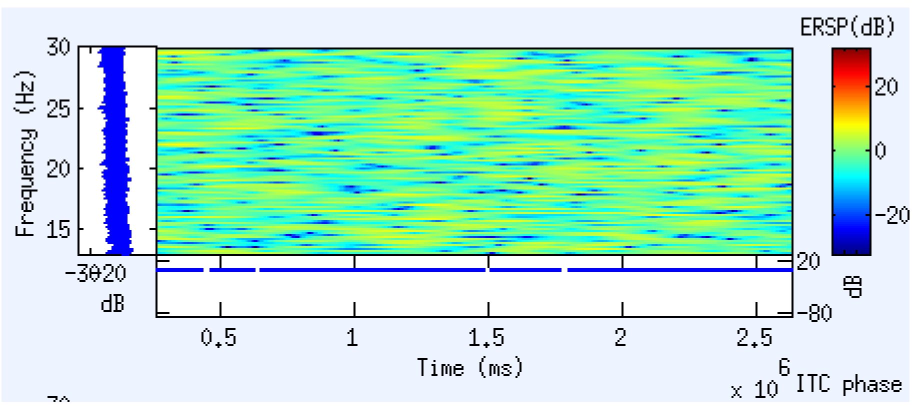


Figure 157: Participant 10 EEG Channel 46 Beta Frequency

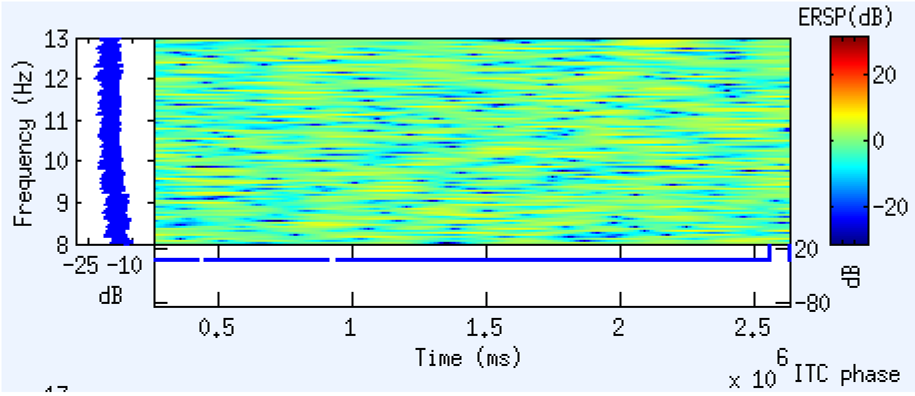


Figure 158: Participant 10 EEG Channel 47 Alpha Frequency

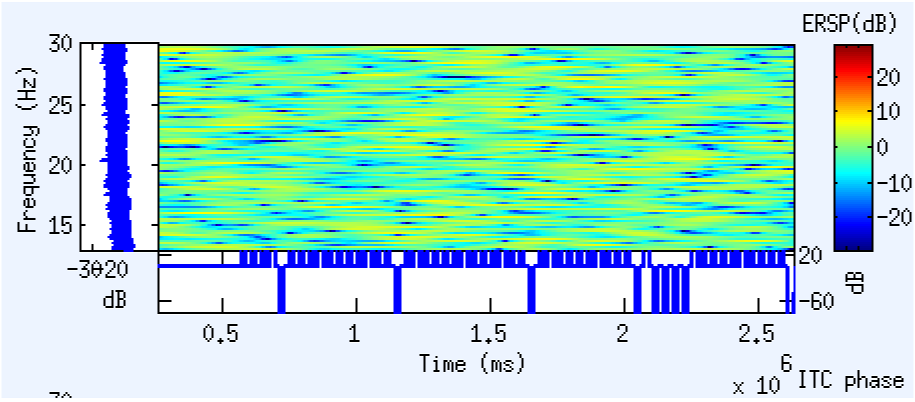


Figure 159: Participant 10 EEG Channel 47 Beta Frequency

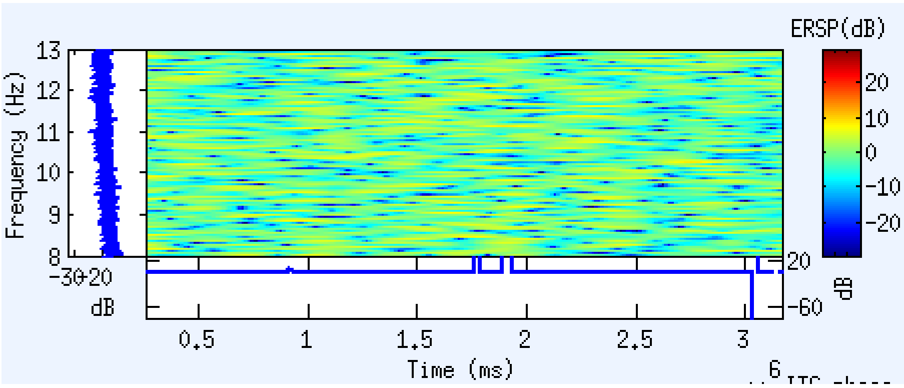


Figure 160: Participant 11 EEG Channel 4 Alpha Frequency

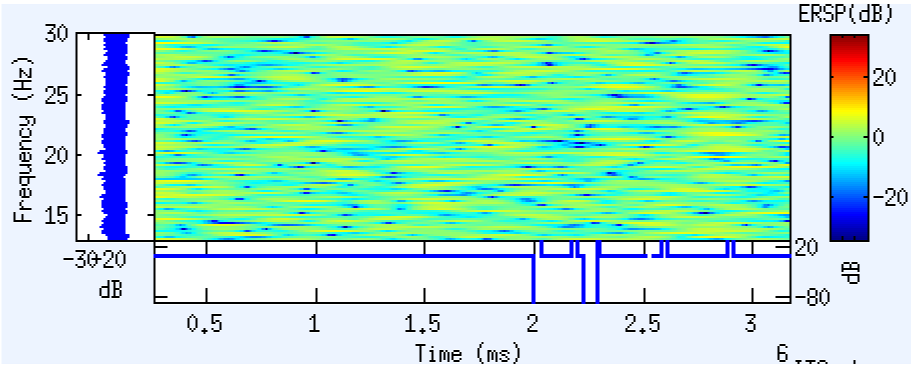


Figure 161: Participant 11 EEG Channel 4 Beta Frequency

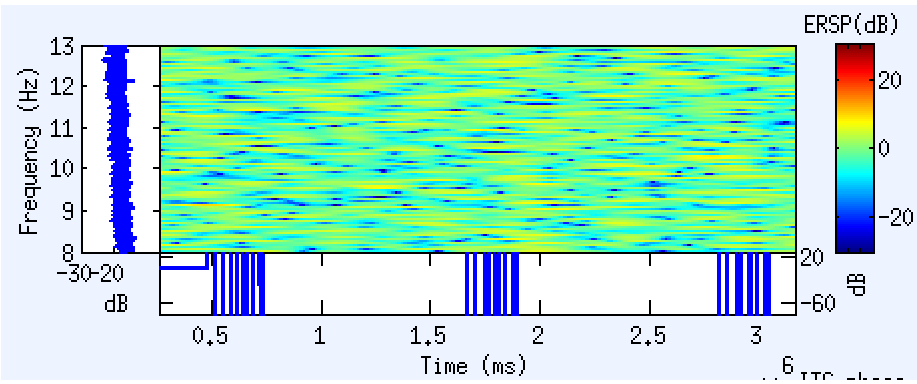


Figure 162: Participant 11 EEG Channel 11 Alpha Frequency

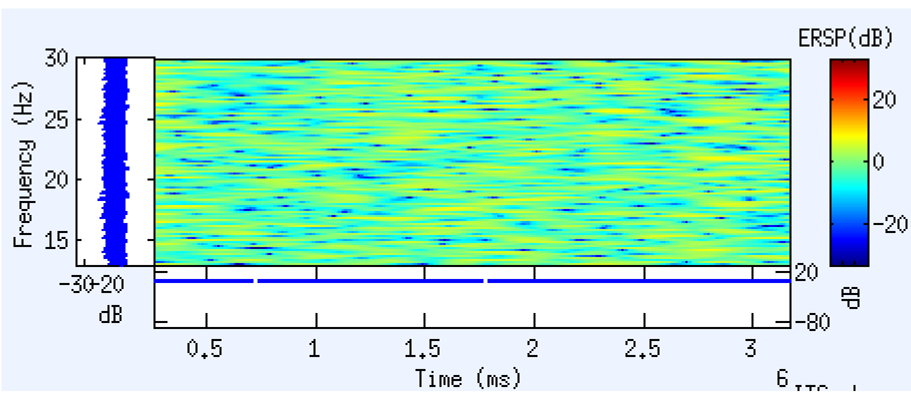


Figure 163: Participant 11 EEG Channel 11 Beta Frequency

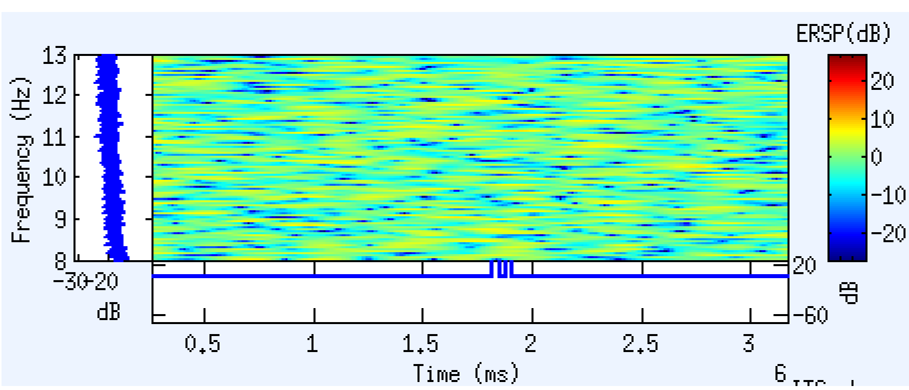


Figure 164: Participant 11 EEG Channel 38 Alpha Frequency

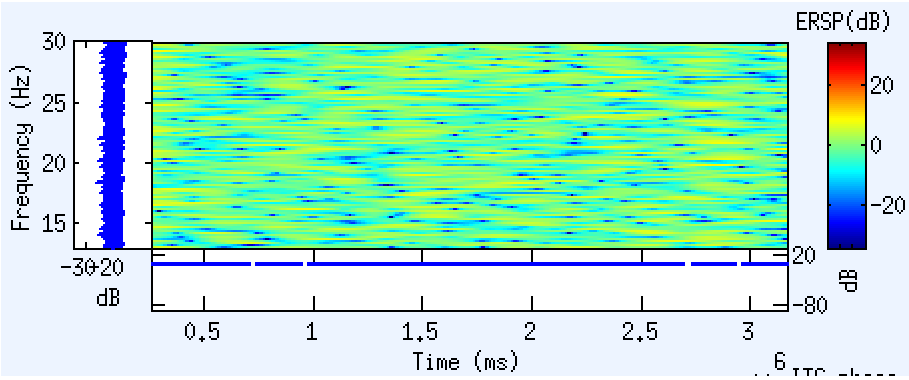


Figure 165: Participant 11 EEG Channel 38 Beta Frequency

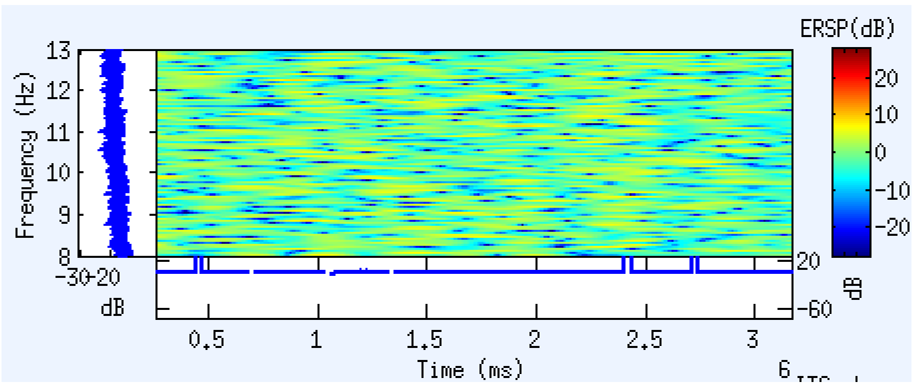


Figure 166: Participant 11 EEG Channel 39 Alpha Frequency

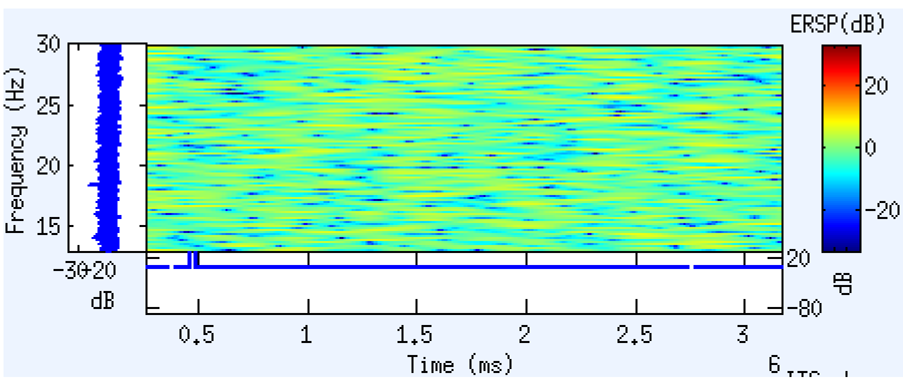


Figure 167: Participant 11 EEG Channel 39 Beta Frequency

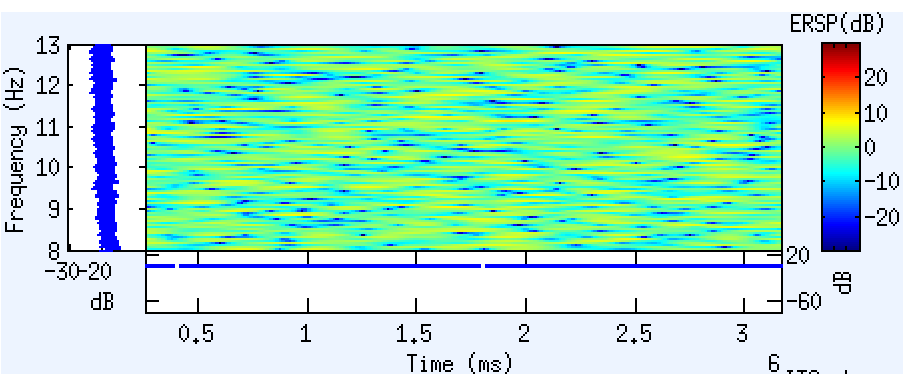


Figure 168: Participant 11 EEG Channel 46 Alpha Frequency

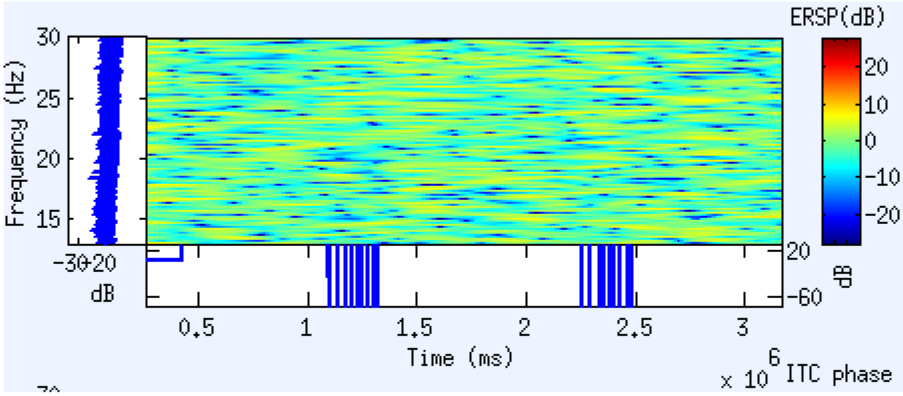


Figure 169: Participant 11 EEG Channel 46 Beta Frequency

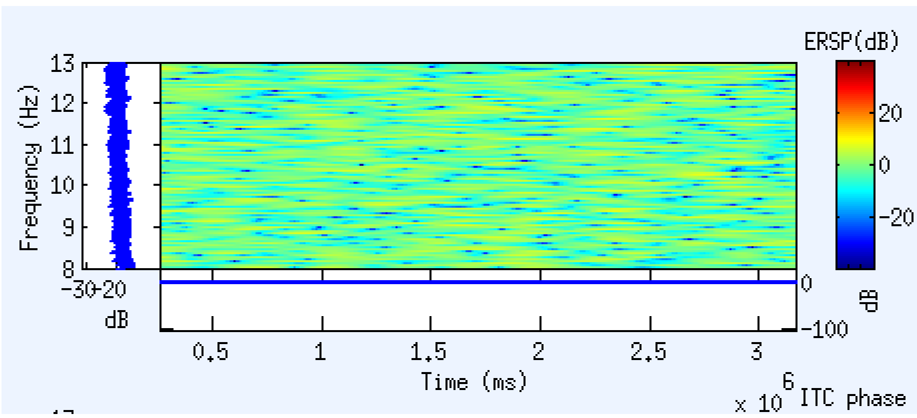


Figure 170: Participant 11 EEG Channel 47 Alpha Frequency

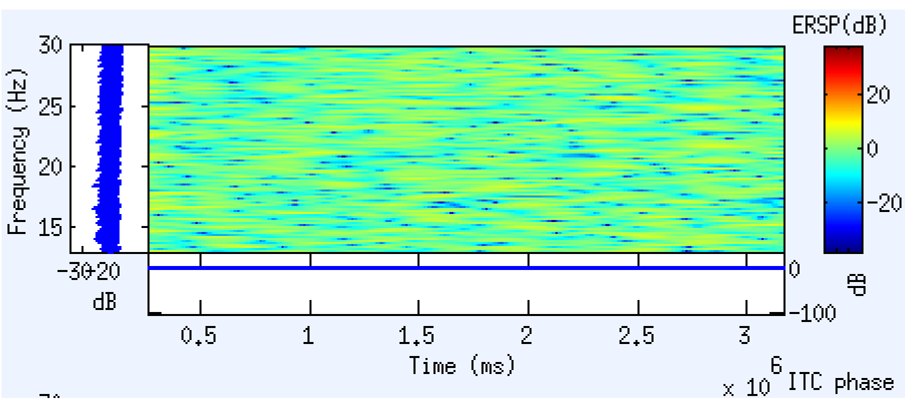


Figure 171: Participant 11 EEG Channel 47 Beta Frequency

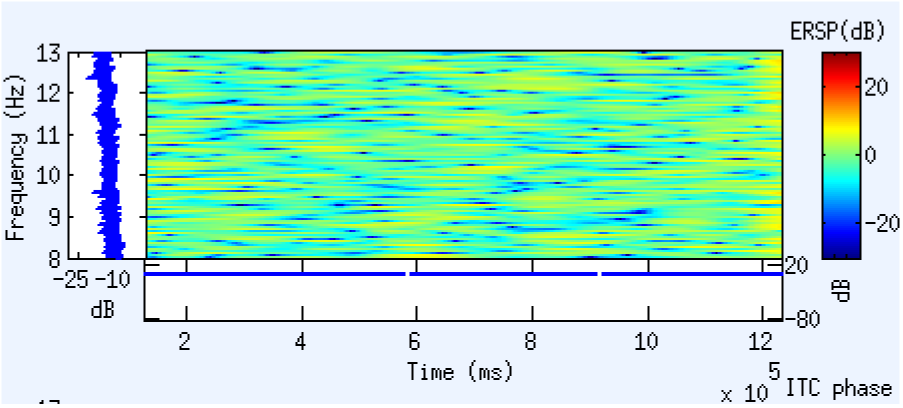


Figure 172: Participant 12 EEG Channel 4 Alpha Frequency

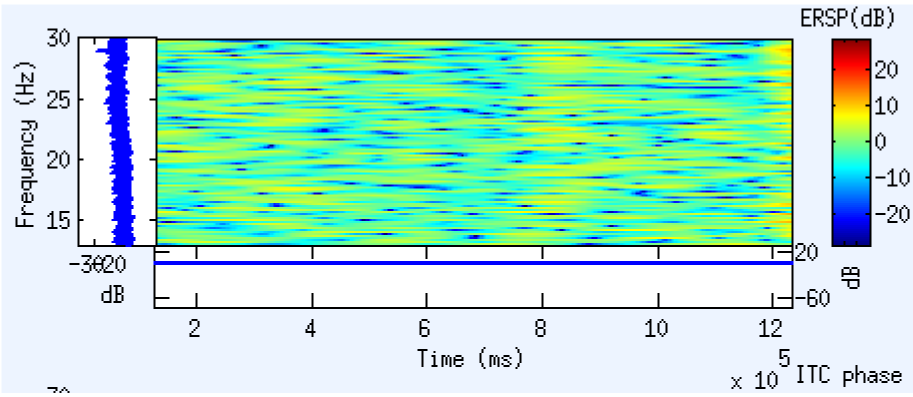


Figure 173: Participant 12 EEG Channel 4 Beta Frequency

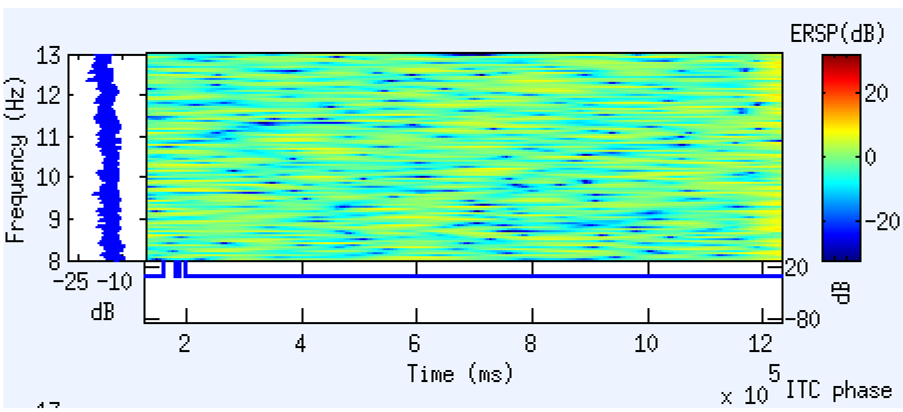


Figure 174: Participant 12 EEG Channel 11 Alpha Frequency

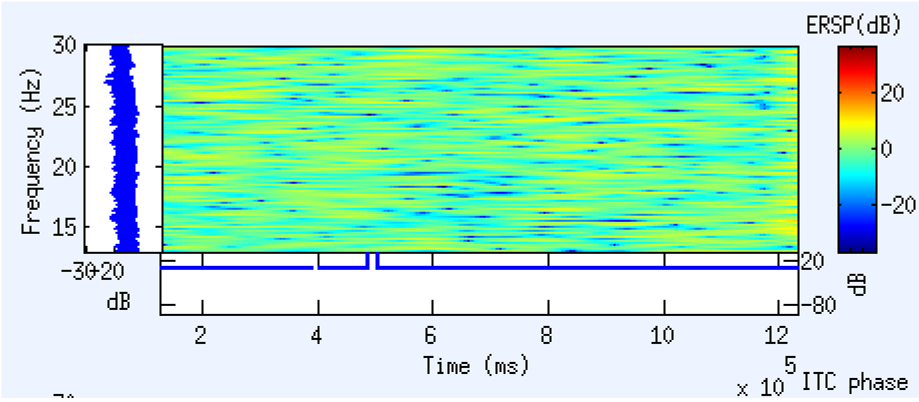


Figure 175: Participant 12 EEG Channel 11 Beta Frequency

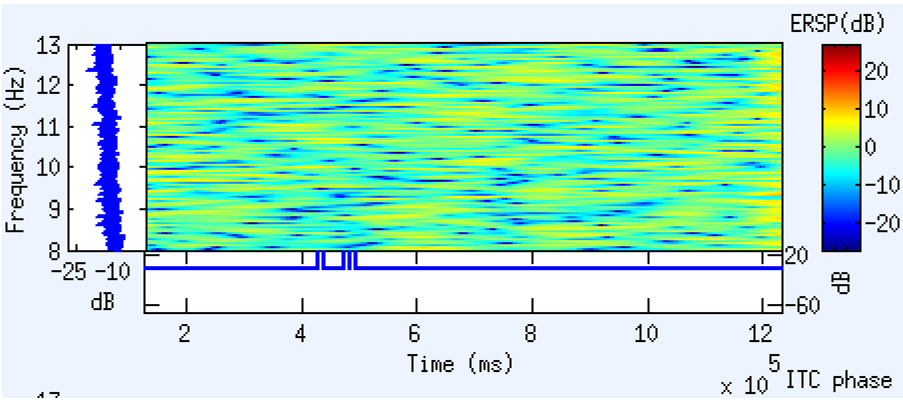


Figure 176: Participant 12 EEG Channel 38 Alpha Frequency

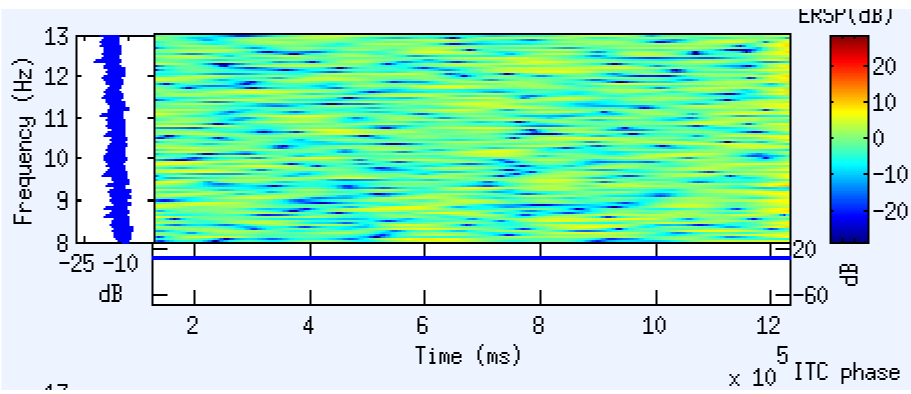


Figure 177: Participant 12 EEG Channel 39 Alpha Frequency

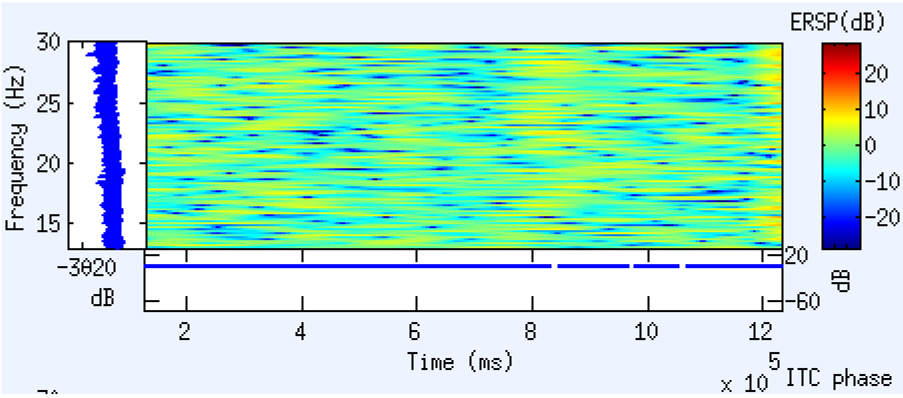


Figure 178: Participant 12 EEG Channel 38 Beta Frequency

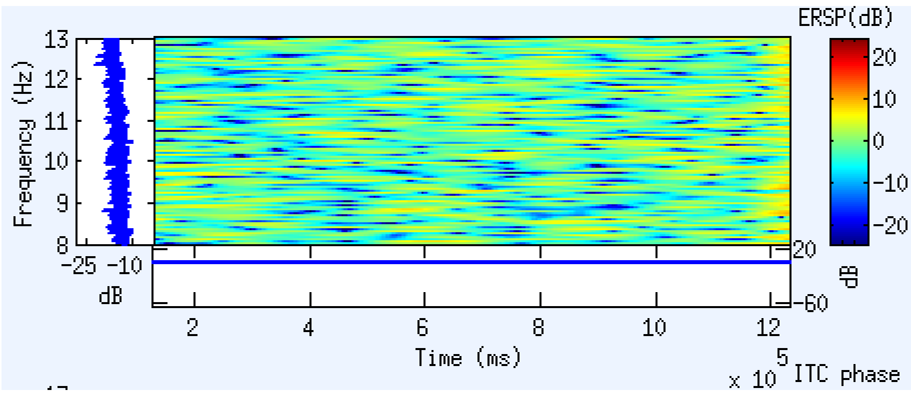


Figure 179: Participant 12 EEG Channel 46 Alpha Frequency

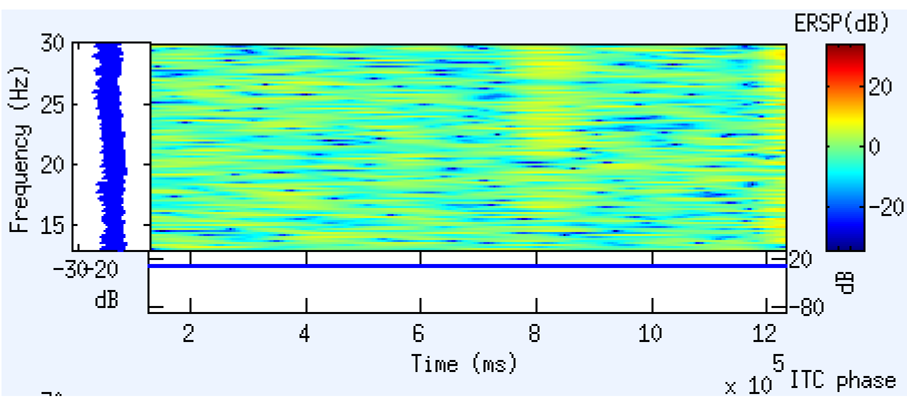


Figure 180: Participant 12 EEG Channel 46 Beta Frequency

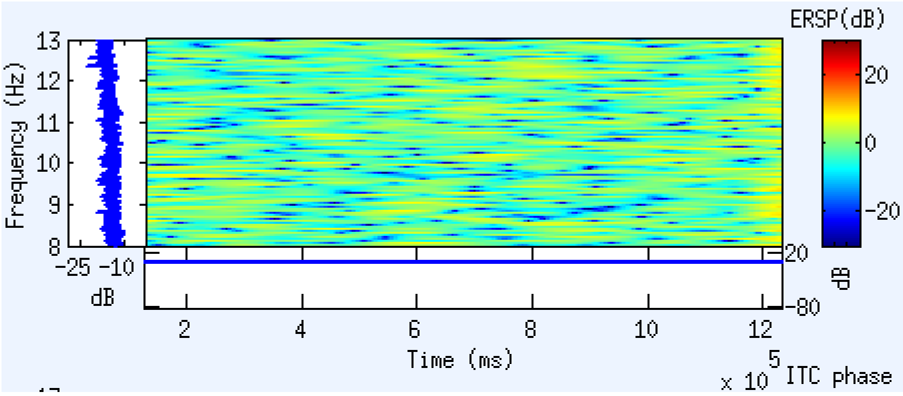


Figure 181: Participant 12 EEG Channel 47 Alpha Frequency

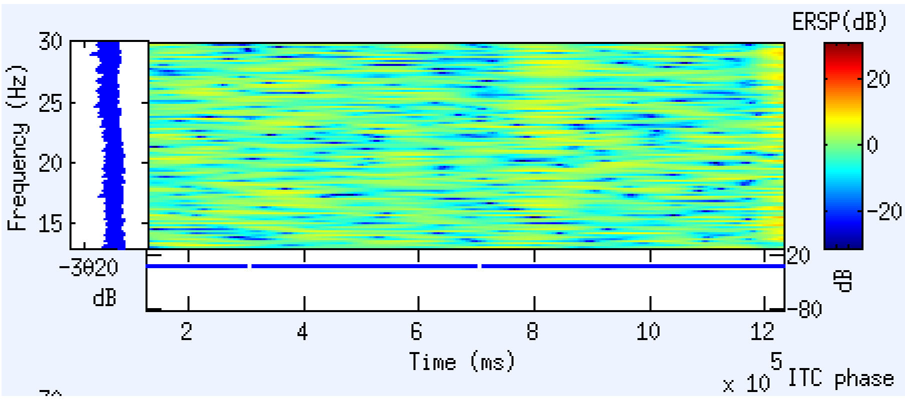


Figure 182: Participant 12 EEG Channel 47 Beta Frequency

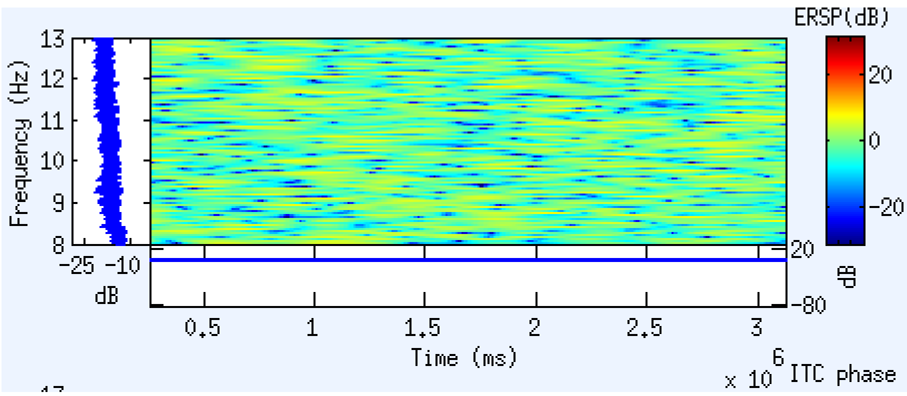


Figure 183: Participant 13 EEG Channel 4 Alpha Frequency

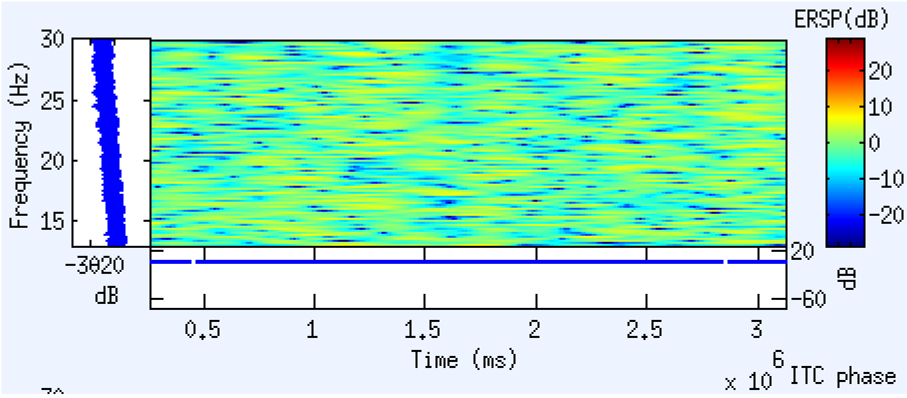


Figure 184: Participant 13 EEG Channel 4 Beta Frequency

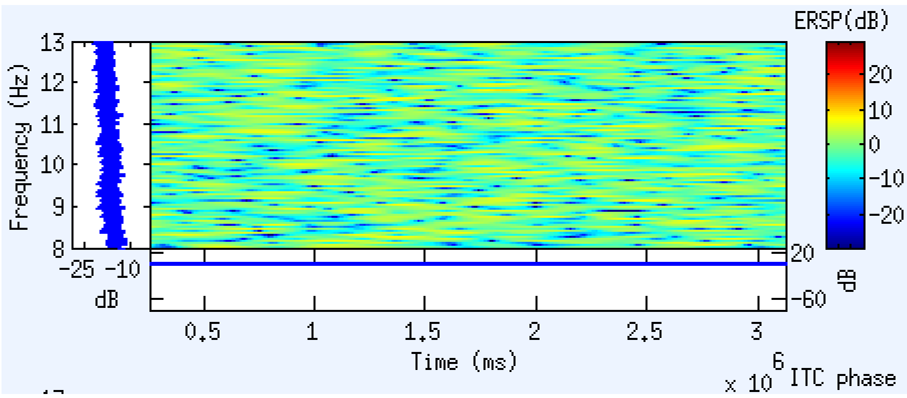


Figure 185: Participant 13 EEG Channel 11 Alpha Frequency

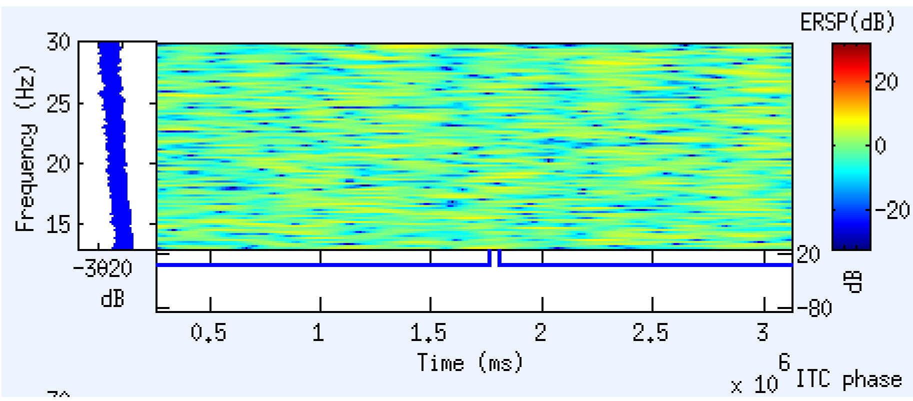


Figure 186: Participant 13 EEG Channel 11 Beta Frequency

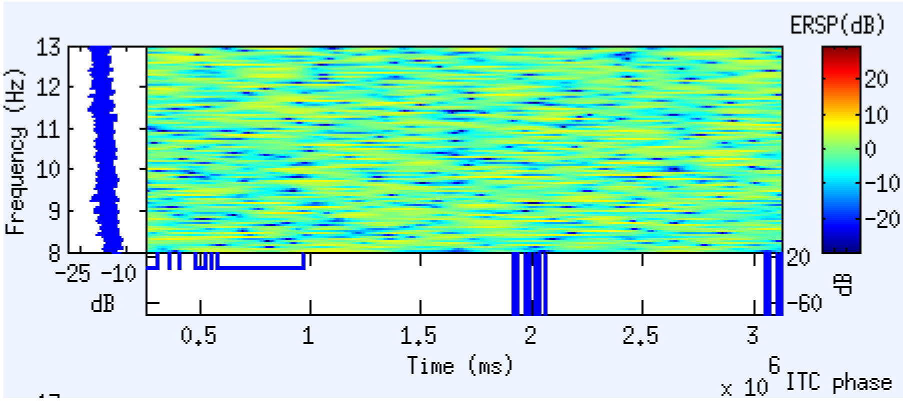


Figure 187: Participant 13 EEG Channel 38 Alpha Frequency

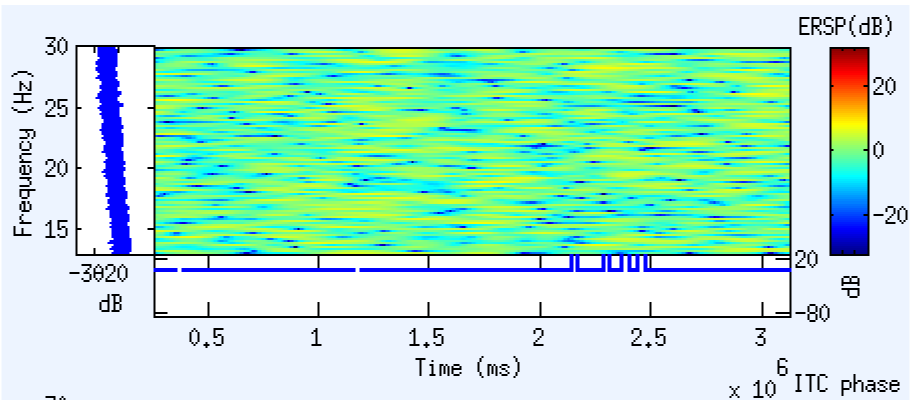


Figure 188: Participant 13 EEG Channel 38 Beta Frequency

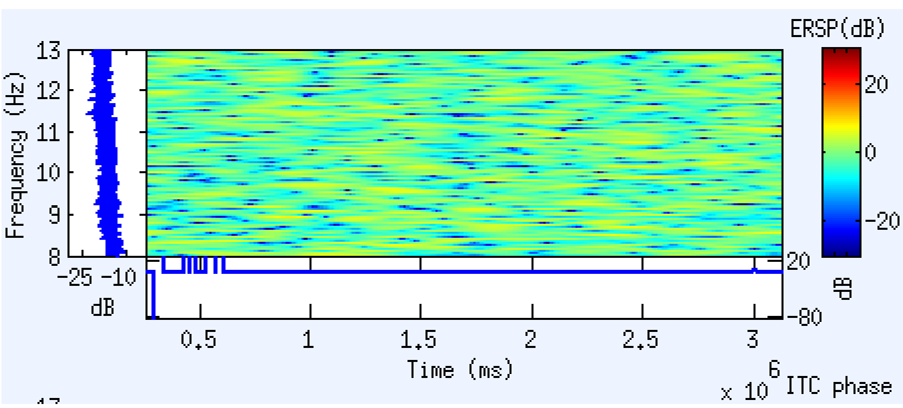


Figure 189: Participant 13 EEG Channel 39 Alpha Frequency

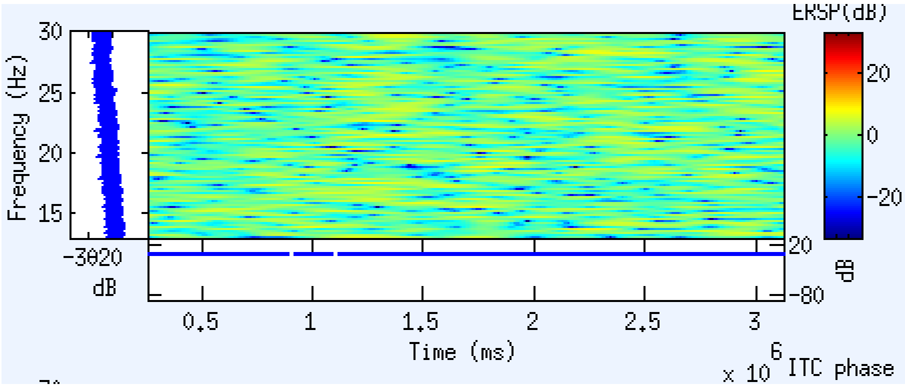


Figure 190: Participant 13 EEG Channel 39 Beta Frequency

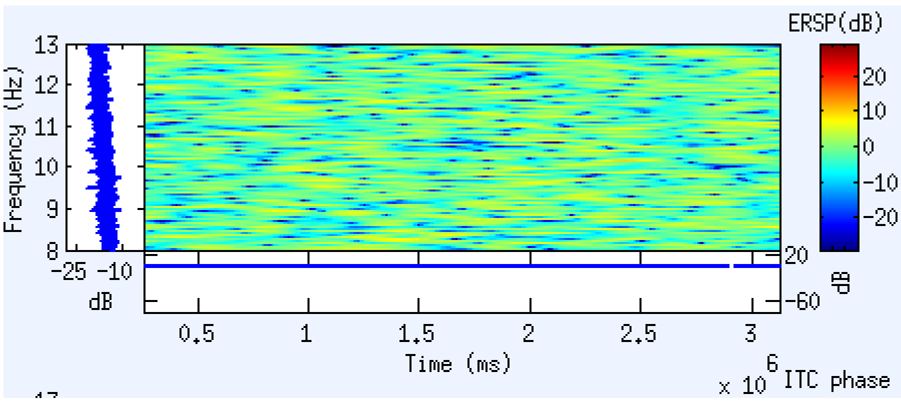


Figure 191: Participant 13 EEG Channel 46 Alpha Frequency

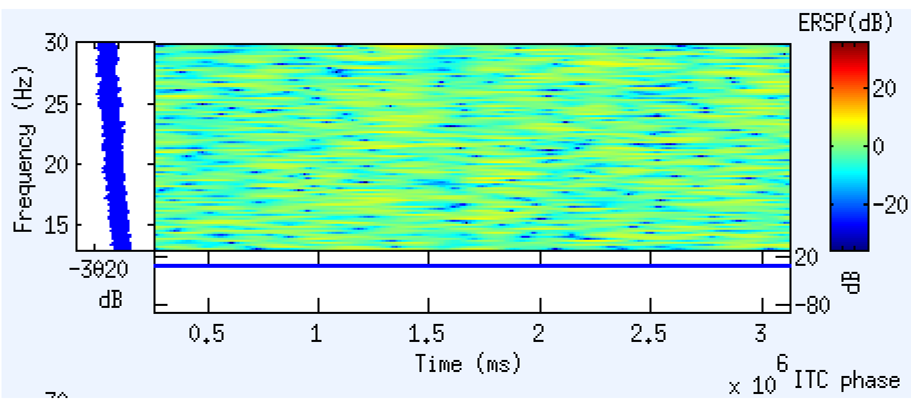


Figure 192: Participant 13 EEG Channel 46 Beta Frequency

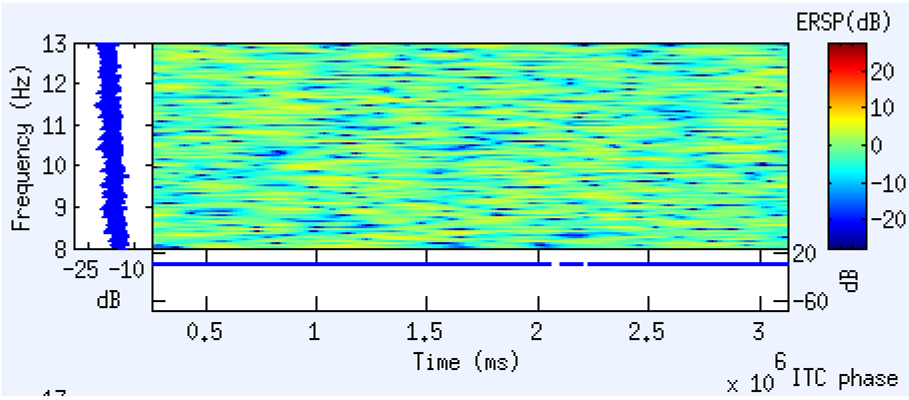


Figure 193: Participant 13 EEG Channel 47 Alpha Frequency

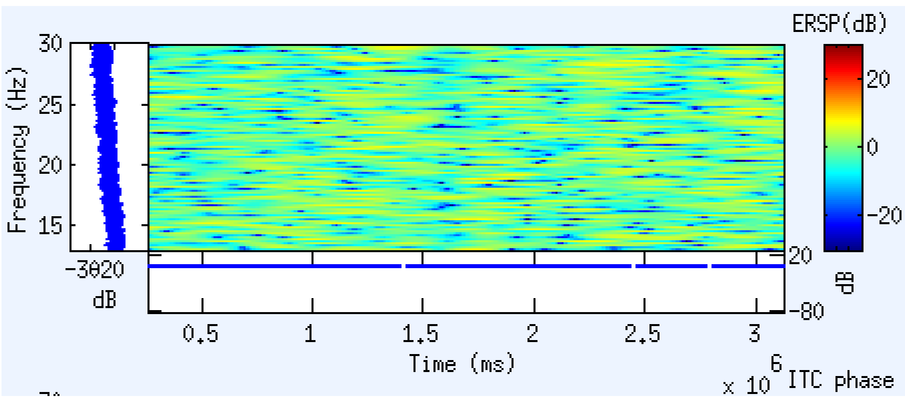


Figure 194: Participant 13 EEG Channel 47 Beta Frequency

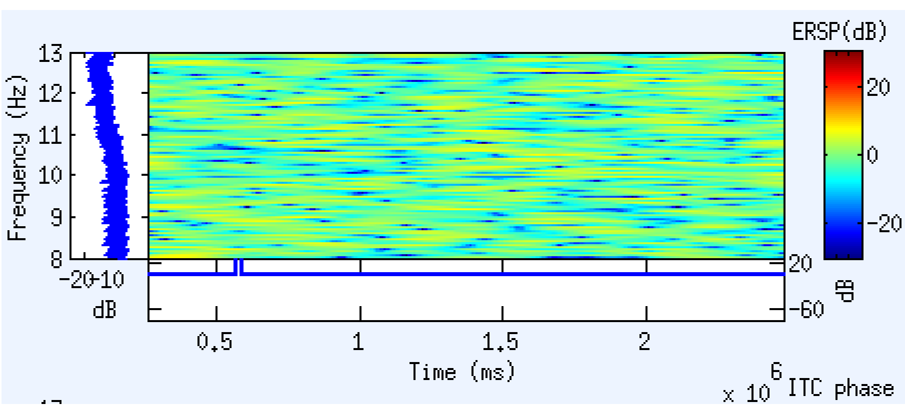


Figure 195: Participant 14 EEG Channel 4 Alpha Frequency

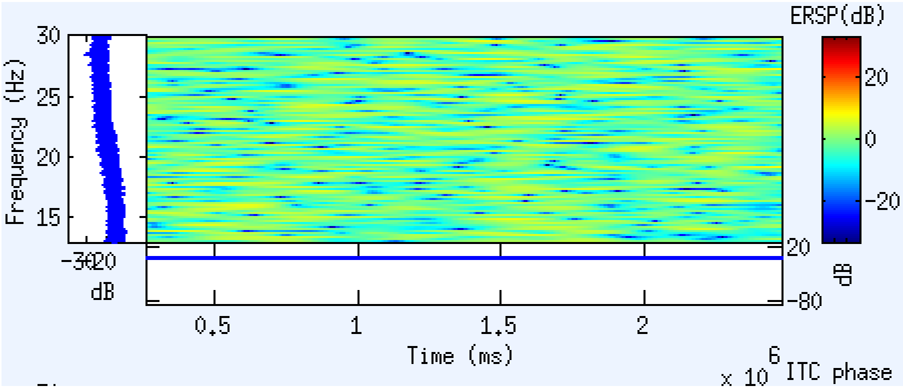


Figure 196: Participant 14 EEG Channel 4 Beta Frequency

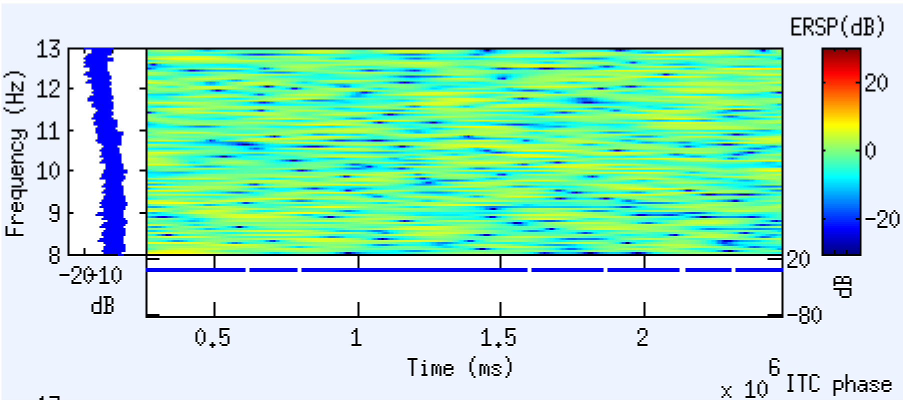


Figure 197: Participant 14 EEG Channel 11 Alpha Frequency

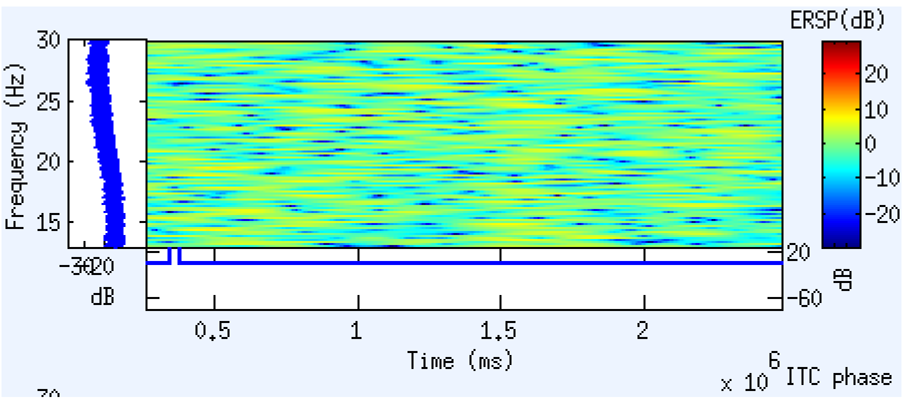


Figure 198: Participant 14 EEG Channel 11 Beta Frequency

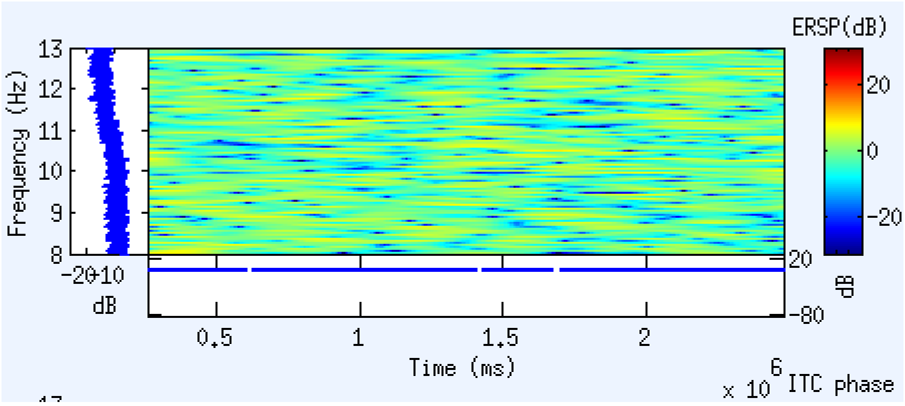


Figure 199: Participant 14 EEG Channel 38 Alpha Frequency

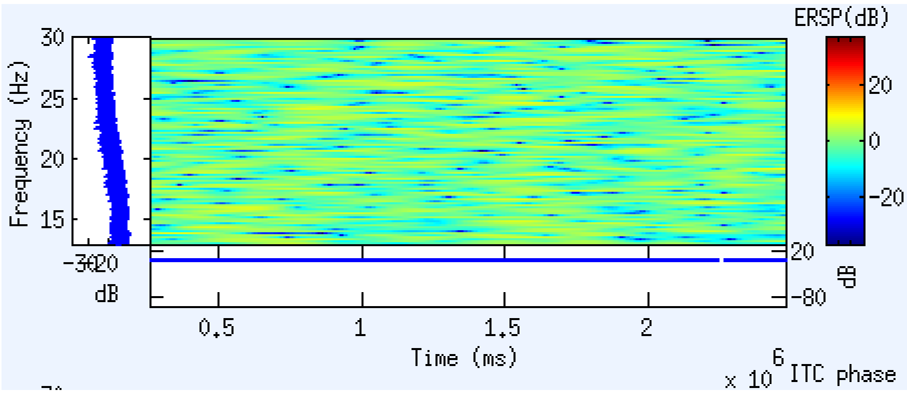


Figure 200: Participant 14 EEG Channel 38 Beta Frequency

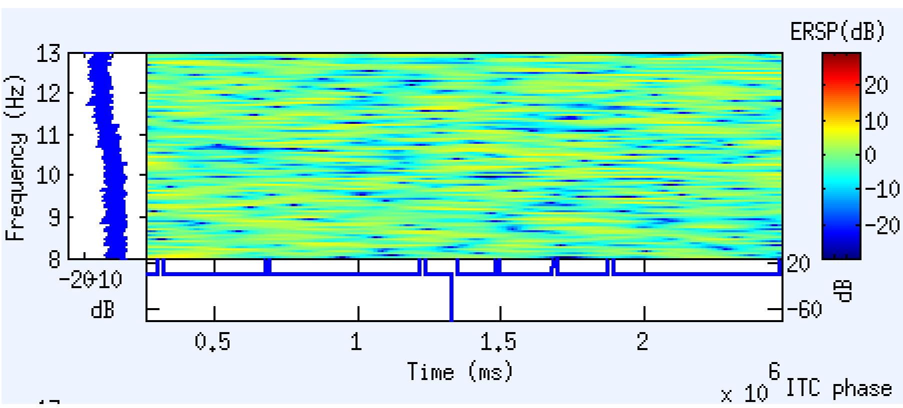


Figure 201: Participant 14 EEG Channel 39 Alpha Frequency

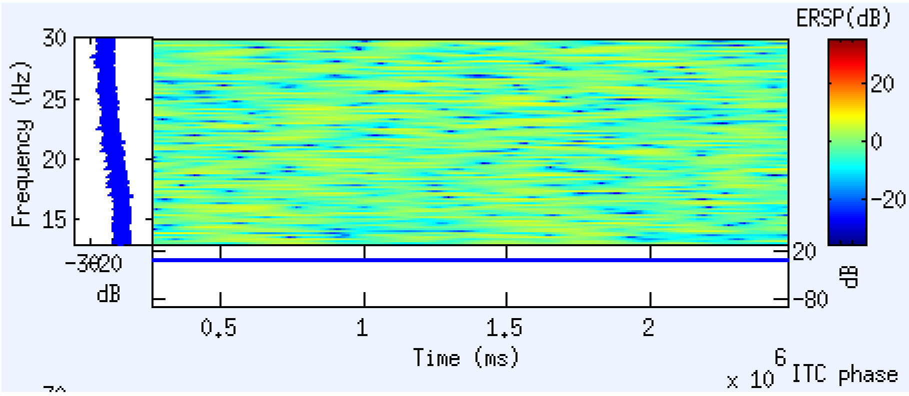


Figure 202: Participant 14 EEG Channel 39 Beta Frequency

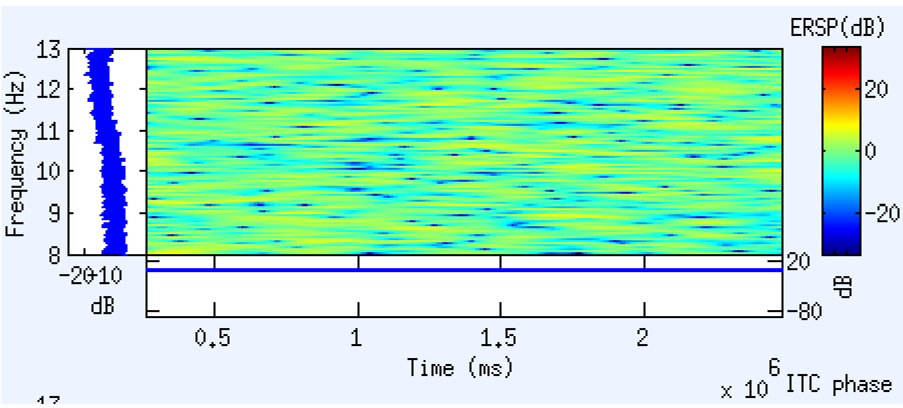


Figure 203: Participant 14 EEG Channel 46 Alpha Frequency

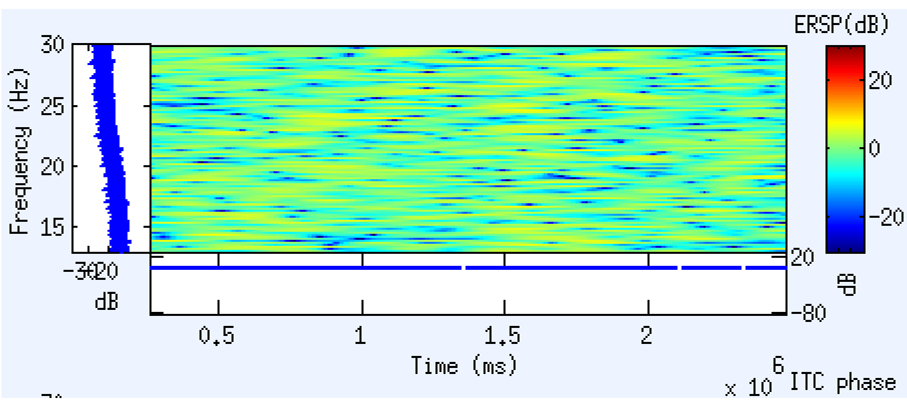


Figure 204: Participant 14 EEG Channel 46 Beta Frequency

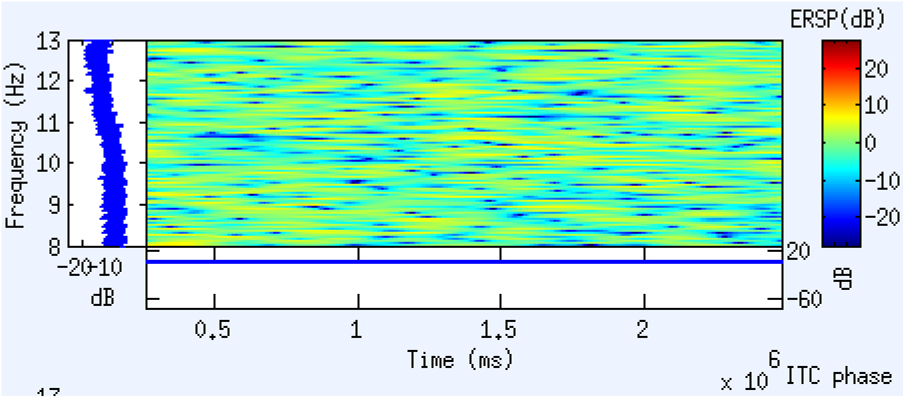


Figure 205: Participant 14 EEG Channel 47 Alpha Frequency

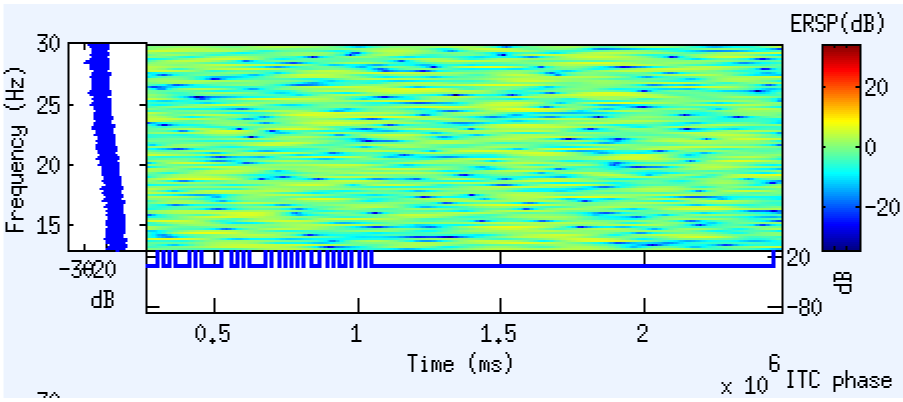


Figure 206: Participant 14 EEG Channel 47 Beta Frequency

Appendix G – Summary of Participant Performance from EEG Data Analysis

	Communications Activity?	Target/Distracter Present?	Participant													
			1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Comms													Wrong turn		
		Distracter		FA		FA		CR	CR					FA		CR
	Comms					Wrong turn										
Segment 1		Distracter		CR			CR	CR		CR					CR	CR
		Distracter		CR				CR						CR	CR	
		Distracter												FA		
		Distracter						CR								
	Comms															
	Comms															
		Target			H		H	H	M				H	H		
		Distracter			CR		CR	CR	CR	CR	CR	CR		CR		CR
		Distracter							CR		CR			CR	CR	CR
		Distracter	CR	FA		FA			FA		FA	CR		FA	FA	FA
	Comms															
		Target	M	H		H			H		H				H	M
	Comms															
	Comms															
		Distracter		CR		CR			CR		CR	CR	CR		CR	CR
	Comms															
	Comms															
		Target	M	M		M			M		M	M	M			
		Distracter				CR				CR	CR		CR			
Segment 2	Comms															
		Target	M	M		M			M	M	M	M	M			
	Comms	Target	H			H			H	H	H	M	H			
	Comms															
		Target	M			H			H	H	H	H	H			
		Target	M			H			H	H	H	H	H			
		Distracter				CR				CR	CR	FA	FA			
	Comms	Distracter				CR				CR	CR	CR	CR			
		Distracter				CR		CR		CR	CR	CR	CR			
		Distracter	CR	CR		CR										
	Comms															
		Distracter														
	Comms	Distracter									CR					
		Distracter														
	Comms															
	Comms															
		Distracter	FA	CR	CR						FA	FA			CR	
		Distracter	FA		CR		CR					FA			CR	
		Target	M	H	H		H				H				H	
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
		Distracter			CR						CR				CR	
		Distracter			CR										CR	
		Distracter			CR						FA				CR	
	Comms															
	Comms															
	Comms															
		Target	M		M					M			M		M	
Segment 3		Target	M		H					H			H		H	
		Distracter								CR			CR		CR	
	Comms															
		Distracter								FA			CR		FA	
		Distracter								CR			FA		CR	
	Comms															
		Distracter								CR			CR		CR	CR
		Target	M							H			H			M
		Target	M				H			H			H			H
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
		Target	M				M			M			H			H
	Comms															
	Comms															
	Comms															
		Distracter	CR				CR		CR	CR			CR		CR	CR
		Target	M				H		M				H		H	H

			Participant													
	Communications Activity?	Target/Distracter Present?	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Comms	Target						M	M		H	H			M	
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
Segment 4	Comms															
	Comms															
	Comms															
		Target						H			M		M		M	
		Distracter											CR		CR	
		Distracter											FA		FA	
		Distracter									CR		FA		CR	FA
		Distracter						CR			CR		CR		CR	CR
		Target						M			M				H	H
	Comms															
		Target						H			H				H	M
	Comms															
		Target						M			M					H
	Comms															
		Distracter														FA
		Target						M								H
	Comms															
	Comms															
	Comms															
		Distracter														
		Target			M			M					H			
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
	Comms															
		Target			H			H								
	Comms															
	Comms															
	Comms															