A Study on
The Use of Augmented Reality and Brain-Computer Interfaces for Improving Situational Awareness

by

Alexander George Minton

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Computer Systems

under the supervision of

Chin-Teng Lin and Hsiang-Ting (Tim) Chen

to

School of Computer Science
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

January 2025

Certificate of Original Authorship

I, *Alexander George Minton*, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Computer Science*, *Faculty of Engineering and IT*, at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Signature: ______________________

Alexander Minton

Date: 15th January, 2025

Place: Sydney, Australia

Abstract

Situational Awareness (SA) forms a core component of life, in both mundane situations to having a critical importance in situations where safety is paramount. To this end, the field of Human Factors has explored the concept of maintaining adequate SA extensively, providing a number of metrics for measurement, as well as safety techniques. However, with the adoption of digital technology, the ability to further augment operator capability has become possible through digital support systems. However, the effects of physiological factors, as well as the extent to which these systems can be developed, have yet to be explored. Additionally, as technology has developed, so too has the ability to integrate the state of the user into worn technology - However, research into integration of this data with worn systems is still in its infancy, with hardware limited to laboratory settings, and the data collected from a wide range of sources, especially in regards to data collected from the brain. To this end, two major gaps can be identified, which this thesis aims to answer - The effects of physiological factors on SA support systems, as well as identification of the specifics of comprehension within the brain, to allow future systems to target this specific area for data collection.

To contribute to this critical field, this thesis aims to investigate three major aspects of these support systems, in order to provide insight into these two gaps. Firstly, the thesis aims to investigate the effects of stress on a Heads-Up Display (HUD)-based Perception Support System (PSS) for auditory stimuli. Secondly, the thesis aims to investigate the effects of circadian fatigue on the same HUD-based PSS. These two chapters (Chapters 4 & 5) aim to identify the effects of the two most common physiological factors on HUD. Finally, this thesis aims to identify the extent to which these systems can be developed through the integration of Brain-Computer Interfaces (BCIs) to further support Human-Machine Teaming. In order to do this, the thesis presents a chapter on the detection and classification of comprehension within the brain with an EEG-based BCI.

This thesis presents three experiments to deepen the understanding of the overall concept of SA. In the first experiment, participants (N = 20) were placed into two virtual reality environments which were developed in the Unity Engine and set in an urban environment and made use of an HTC Vive Eye Pro Head-Mounted-Display. These environments were designed to induce different stress levels, and participants were required to complete an audio and audio-visual perception task with and without HUD support, whilst maintaining awareness of the virtual environment. In the second experiment, participants (N = 19) were once again placed into the same virtual environment as the first experiment, however their fatigue level was predicted using machine learning techniques based on hours slept, using the Fatigue Impairment Prediction Suite. In the third experiment participants (N = 19) were presented with a tool, and a context in which the tool had to be used, and were required to identify if this would be detrimental, helpful, or irrelevant to successful task completion, while EEG data was collected in order to isolate and classify their selected categorization.

The results of these studies indicate that current HUD-based support systems continue to function effectively under physiological load, as well as improve performance in certain aspects, indicating suitability for use in situations with high physiological load. This finding demonstrates significant potential for application of HUD-based systems in fields where situational awareness can be affected by physiological factors, providing an avenue of increasing operator awareness and safety. Additionally, the results of the EEG-based study demonstrate the ability to isolate and classify the brain dynamics of comprehension, in order to allow machine agents to support the user in their comprehension of a situation. This EEG biomarker allows exploration into its uses for machine classification, potentially improving classification accuracy, opening future research into EEG-based situational awareness support systems.

DEDICATION

To the giants whose shoulders I have stood upon, and to those who try to make the world a better place for all. May this work help those in the future reach even further heights...

Acknowledgments

First and foremost, I would like to thank Dist. Prof. Chin-Teng Lin and Dr Hsiang-Ting (Tim) Chen. Without their guidance, patience, and supervision, this thesis would not have been possible. Thank you for the continuous support in such a multi-discipline thesis, and helping me with all of the skills required to interact with each research area.

I would also like to thank all of my colleagues in the Computational Intelligence and Brain Computer Interface Group and the Human-Centred AI Centre for the help, discussion, and pointers on how to approach problems and data analysis. Special thanks to Howe, Thomas, and Daniel for guiding me through the EEG aspect especially.

I am extremely grateful to the talented undergraduate students I had the pleasure of working with, who assisted me with data collection for the EEG pilot and study. Without their help both when I was based overseas due to COVID-19, and when I was using a large number of EEG channels to record, the experiments would not have been able to be completed in a timely manner. Many thanks to Leesandra and Fadi for their dedication.

I would also like to acknowledge the staff members that allowed me to engage with the community and teach at UTS, Dr Jaime Garcia Marin and Dr Avinash Singh, as without their support I would not have discovered the joy of teaching at University.

I would like to acknowledge the University of Technology Sydney School of Computer Science and UTS Techlab for the facilities and support, as without their support of excellent spaces and equipment, the experiments in this thesis would not have been possible. I am also extremely grateful for the financial support provided by the International Research Scholarship and the FEIT Scholarship, as this allowed me to complete this research and contribute to the wider academic community.

Lastly, I would like to thank my friends and family for their support throughout my entire thesis, as without all of you, this thesis would not be completed. To my friends in the UK, Ireland and Denmark, who number too many to list in this one page, I thank you all for your tremendous support throughout all the years. To my newfound friends in Australia, especially Carlos, Aden, and Claudine, thank you for welcoming me to this country and supporting me when I needed it most. To my parents, Freda and David, and my brother, Chris, thank you for all of your support, guidance, love, and patience all of these years.

Without everyone listed above, and those who I wasn't able to fit on this page but still supported me, this PhD journey would not have been possible, and I am extremely grateful for each and every one of you.

List of Publications

Related to the Thesis :

1. Minton, A.G. et al. 2025, "A Longitudinal Study on The Effects of Circadian Fatigue on Sound Source Identification and Localization using Heads-Up Displays" *Accepted for publication in 2025 ACM Special Interest Group on Computer-Human Interaction. ACM.*

2. Tian Y. , Minton, A.G. et al. 2021, "A Comparison of Common Video Game versus Real-world Heads-up-display Designs for the purpose of Target Localization and Identification" *In 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 228-233). IEEE.*

3. Minton, A.G. et al. 2024, "An Investigation into using Heads-Up-Displays to Improve Non-Line of Sight Spatial Audio Source Detection & Localisation under Real-World Stress Factors" *being prepared for publication in 2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*

4. Minton, A.G. et al. 2024, "Identifying a Biomarker for the Brain Dynamics of Contextual Semantic Categorization using Electroencephalography" *being prepared for publication in IEEE Transactions on Cybernetics*

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 The Concept of Situational Awareness

Situation Awareness forms a core part of our lives, ranging from the mundane, where simply being aware of the current time when you have a meeting in a few hours helps to ensure you won't miss it, to being aware of a plane entering a stall, where failure to quickly take the correct actions can result in disaster. To increase our situation awareness, we have created a wide variety of aides [1, 2], such as simple calendar alarms reminding us before meetings, to the voice warning systems used in aircraft to alert pilots to hazards. However, state-of-the-art systems fail to take a user's understanding into account, which can lead to failures, which in stressful situations can quite often be fatal. To improve this human-AI teaming, recent studies have integrated BCIs with SA support systems [3], to some success. However, this model was based on a user's reaction as to whether an alert was task-relevant or not, rather that specifically identifying their understanding of the alert. By identifying their specific understanding, we can ensure that they correctly understand the meaning of a task-relevant alert, rather than just that the alert itself is task-relevant, as if they take the wrong action based on a misunderstanding, that could potentially be more dangerous than not understanding an action needed to be taken in the first place.

1.1.1 Failures in Situational Awareness

The most widely regarded model for situation awareness, the Three Level Model by Endsley [4], identifies three main aspects of situation awareness: Perception of the environment, Comprehension of the information perceived in relation to the situation as a whole, and Projection of how the situation will evolve over time. From these aspects, an individual can come to a decision about what action they wish to take to affect the situation. While there are multiple models that attempt to define such an abstract concept as situation awareness, other widely regarded models also highlight the importance of comprehension, such as Bedny and Meister's Interactive Subsystems model [5], where the first step of orientation in a situation is determining the meaning of perceived information, or Smith and Hancock's Perceptual Cycle model [6], where the available information modifies an individual's knowledge of the current situation. From these models, we can see that comprehending a situation is key to maintaining good SA. However, just as comprehension is key to good SA, a failure in comprehension can prove to be devastating to the appropriate handling of a situation. Perception of snowy weather when you are intending to drive to your meeting is important, but unless you comprehend that the drive through the snow will be slower, you will not project that it will take longer than usual to drive to the office, thus potentially missing a meeting. This remains true in even more dangerous situations such as aircraft, where even if you perceive your airspeed dropping and a lack of response in the aircraft controls, unless you comprehend that these are potentially the signs of a stall beginning you are not able to identify the need to take corrective action to recover from the stall, leading to a deterioration in the situation where it is not until the aircraft begins losing altitude that the pilot realises, or worse - the pilot does not realise at all.

The importance of comprehension in maintaining good SA can clearly be seen, as Endsley and Jones identify that 'errant mental models' are a major cause of failures in SA [7]. However, whilst it seems logical in this case to allow automated systems to handle matters when able, as automated systems are not at risk to as many of the factors of SA as humans, they themselves cause issues when a human is required to interact with them, causing issues which Endsley and Jones call 'out-of-the-loop syndrome' [7]. This is where the actions taken by the system are unknown, or what actions the system is supposed to handle are unknown, causing a lack of understanding in how the situation is being dealt with. Furthermore, an over-reliance on automation has been shown to have a negative impact on performance when unaided [8, 9], indicating a requirement to provide support only when necessary. However, to do this is no easy task, as whilst it is easy to instruct a pilot to 'pull up' with a voice warning system, or for a semi-autonomous car to brake if appropriate, these systems do not take an operator's comprehension into account, merely skipping to the Decision or Action step (if referencing Endsley's Three Level Model), bypassing an individual's SA entirely.

The greatest difficulty comes in identifying or measuring SA as a whole, as such an abstract and complex measure is often hard to quantify. Efforts have been made to provide metrics to measure it, though it is regarded as a qualitative metric, with quantitative metrics being applicable only when the task itself is directly tied to an individual's SA, identifying their SA through the actions they take to perform the task [10]. In more complex tasks, the task is often paused and the participant is subjected to SAGAT [11], a questionnaire style designed to identify a participant's SA based on the task they are currently performing, by querying them about the current situation, what it means, and what the situation will look like in the future. However, these methods have major issues in their feasibility in real-world applications. Performance-based SA measurement requires an individual to take an action, or begin to take an action, before it can be determined whether their comprehension is correct by whether the taken action is appropriate, and SAGAT requires, quite literally, time to be 'paused' during a simulated situation so that a clearer understanding of an individual's comprehension of a situation can be ascertained. This demonstrates a clear need for a real-time system that can identify an individual's comprehension of a situation without having to rely on performance or user action as an indicator.

The use of BCIs with relation to SA has grown greatly in the recent years, both due to the recognition of the importance of SA in stressful tasks [12], as well as the growing complexity of said tasks over time [13]. As interruptive tasks such as SAGAT cannot be done in real-time, nor are they feasible in a real-world environment, another way of detecting, and potentially correcting, SA must be achieved in order to provide real-time assistance. A recent study by [3] revealed that when a passive BCI was integrated into cognitive modelling of aircraft pilot's responses to alerts and messages, the classification accuracy rose to 87% from 72% with the model that did not consider the EEG data. This highlights the importance of integrating neurological data into current SA support systems, as a higher success rate in classification of response, as well as the ability to rapidly respond, can help prevent accidents from occurring. To this end, the integration of an individual's comprehension of the situation and the meaning of elements within it into SA support systems serves as a clear method of increasing the overall effectiveness of SA support systems.

1.1.2 Current Methods of Improving Situational Awareness

1.2 Thesis Structure

1.2.1 Thesis Definition

The mechanism of Situational Awareness is exceptionally broad, ranging from perceiving information in the surrounding environment, to comprehending the meaning of this information, to projecting how the environment will change without interference, to the specific interference needed to achieve desired goals. As this broad field is too large for a single thesis to explore thoroughly, this thesis will primarily focus on the investigation of improving Perception of stimuli through Heads-up-displays, as well as determining the feasibility of improving Comprehension through the detection of related brain dynamics with EEG. In this thesis, Virtual Reality (VR) has been used to emulate a naturalistic environment for the purpose of measuring SA, with the Heads-Up-Display overlaid over the view to simulate a wearable HUD system.

1.2.2 Aims, Objectives, & Research Questions

The overarching goal of this thesis is to identify the key aspects of Situational Awareness (SA) that can be augmented by human-support systems, how environmental factors affect an individual's SA, and the integration of Brain-Computer Interfaces (BCIs) into a support system to further improve human situational awareness.

1.2.2.1 Aims

The aims of this thesis relate to the augmentation of SA under Endsley's Three-Level model through the use of a Heads-up Display (HUD). As part of this, it must be identified that any potential solution does not become a liability when the user is under the pressure of environmental factors. In addition, the integration of BCIs may potentially allow the system to provide support in the Comprehension level, however this capability must be determined. To this end, the aims of this thesis are as follows:

1. Provide support for the Perception Level of Situational Awareness through the use of a Heads-up Display

2. Ensure the Heads-up Display functions under environmental factors such as stress or fatigue

3. Provide support for the Comprehension Level of Situational Awareness

1.2.2.2 Objectives

To derive objectives from the aims stated above, key points can be drawn to identifying the interaction between HUD use and environmental factors, as well as the requirement to identify the brain dynamics behind the process of comprehension. To this end, the objectives of this thesis are as follows:

1. Identify the effects of Stress on perception capabilities and the use of Heads-up Displays
2. Identify the effects of Fatigue on perception capabilities and the use of Heads-up Displays
3. Identify the brain dynamics behind comprehension level SA, and provide a machine-learning system to classify user comprehension

1.2.2.3 Research Questions

To further deconstruct the objectives into research questions, it is possible to identify two main factors to investigate, these being the effects or processes in each objective, and then the effectiveness of a support system for each factor. For stress and fatigue, the effects on perception must be identified, and then the continued effectiveness of the support system must be identified. For comprehension support, firstly the brain dynamics of the process must be identified, and then the ability to make use of identified biomarkers must be confirmed in order to highlight application potential. To this end, the research questions are as follows:

1. Effects of Stress on SA:
   a) How does stress affect the capability to perceive non-visual aspects of the environment
   b) Can a Heads-up display continue to provide tangible benefits when the user is in a state of stress
2. Effects of Fatigue on SA:
   a) How does fatigue affect the capability to perceive non-visual aspects of the environment

   b) Can a Heads-up display continue to provide tangible benefits when the user is in a state of fatigue

3. Comprehension in SA:
   a) Does the process of comprehension in SA have a distinct biomarker within the brain
   b) Is it possible to classify brain signals in order to provide a support system with a classification of user understanding

1.2.3 Stakeholders

The primary stakeholders of this research will be UX designers, specifically those focusing on wearable Heads-up-displays. As wearable XR technology increasingly enters mainstream use, the various environmental and physiological conditions of use will become increasingly important. They will be able to use the results of this research to guide their system design, improving the usability of their systems.

A secondary stakeholder of this research will be researchers in the SA and BCI fields, with the capability to identify the process of comprehension within the brain, in regards to situational awareness. The identification of these processes will allow both development of BCI systems that can make use of these for both research and industrial purposes, as well as furthering the field of SA research by opening the possibility for real-time detection of higher levels of SA.

1.2.4 Approach

The work presented in this thesis utilizes the use of Virtual Reality to emulate environments which may not be suitable for conducting experiments in-situ. In order to make use of this, appropriate naturalistic environments were constructed within the Unity Engine, and then a suitable physical experiment environment was utilized to provide a further closeness to a real-use environment. To provide a high-quality VR view, this thesis makes use of an HTC Vive Pro-Eye VR HMD, and utilizes HTC's Surround Sound Spatialization to ensure compatibility with the HMD's built-in speakers.

The recording of the EEG data is conducted with data quality and fidelity in mind rather than feasibility for real-world deployment, and as such, the EEG data is collected with a Neuroscan 128 channel wet-sensor cap. This data is then processed with state-of-the-art approaches using the EEGLAB toolbox in MATLAB 2021. Some of the processes used for pre-processing include down-sampling, high- and low- band pass filtering, Independent Component Analysis (ICA), dipole source localization, and automatic artefact removal. The processed EEG data is then clustered for all participants to provide a combined dataset for analysis.

In regards to EEG results, the EEG data is primarily investigated in both channel-form, and component-form (from ICA), specifically in Event-Related Potentials (ERPs) and Event-Related Spectral Pertubation (ERSP), to allow for differentiation of the conditions and identification of biomarkers for future usage.

1.2.5 Findings

The results of the first experiment (*Chapter 3*) indicate that prior experience can be successfully leveraged when designing systems with usability in mind. This can guide both system development as well as allowing improved training of users on the system by utilizing prior knowledge.

The results from the second experiment (*Chapter 4*) show that certain levels of stress can improve performance, however this is not universal. This indicates that both stress level as well as the task being performed must be accounted for when designing support systems in order to ensure effectiveness of these systems.

The results from the third experiment (*Chapter 5*) show that fatigue can have a varying effect on performance metrics, demonstrating the importance of consideration as to what metrics the system is attempting to improve. Furthermore, the results show that such systems can have a detrimental effect on certain performance metrics, highlighting the importance of additional support in these areas to account for unintended degradation in performance due to system usage.

The results from the fourth experiment (*Chapter 7*) demonstrate that the process of comprehension can be detected within the brain using current EEG recording equipment and analysis techniques, indicating the potential for a closed-loop system to be developed that make use of these identified features.

1.2.6 Chapter Organisation

This thesis is divided into 8 chapters that detail the introduction, background literature, studies conducted as part of the thesis, overall results and conclusion, and future research application. An adaptation of Endsley's Three Level model (Figure 1.1) has also been provided to highlight how each chapter corresponds to the related area of Situational Awareness.

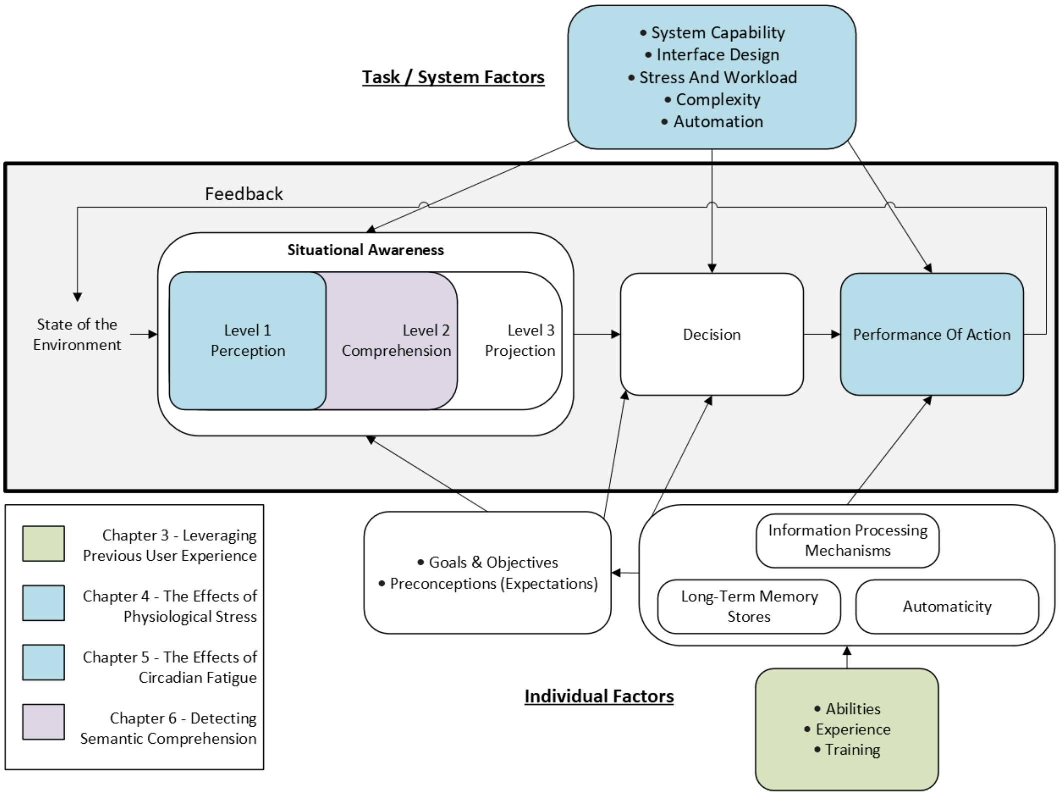


Figure 1.1: An adaption of Endsley's Decision Making Model incorporating the Three Level SA model [4], highlighting the sections explored in this thesis

- Chapter 1, the *'Introduction'*, provides the background motivation for this thesis, as well as briefly describes the concept of Situational Awareness (SA), and the importance of detecting failures in SA. This chapter also defines the structure of the thesis in regards to Aims, Objectives, and Research Questions, as well as outlining the contribution and importance towards potential stakeholders.

- Chapter 2, entitled *'Situational Awareness, Physiological Factors, and Improving Situational Awareness with Heads-Up Displays'*, provides a comprehensive literature review regarding the current state of modelling the concept and process of SA, and the effects of physiological stress and fatigue on SA. Additionally, this chapter also reviews current uses of Heads-Up-Displays as SA aides, to guide development, and finally reviews the current understanding of how the process of comprehension occurs within the brain, to provide a further understanding of how this process can be leveraged for improving SA.

- Chapter 3, entitled *'Leveraging Previous User Experience for Improved User Performance'*, presents a study into whether previous user experience on similar HUD systems can be leveraged to provide tangible benefits for a task and system with which the user has no prior experience. These results provide insight into HUD development, and guide the HUD design for future chapters.

- Chapter 4, entitled *'How Physiological Stress affects HUD use and SA for Target Localization and Identification'*, presents and describes a study on how stress affects HUD usage and SA. The results from this study provide an insight into how stress can affect HUD use, and how SA support systems should account for these effects.

- Chapter 5, entitled *'How Circadian Fatigue affects HUD use and SA for Target Localization and Identification'*, further investigates physiological factors by presenting a longitudinal study on the effects of fatigue on HUD usage and SA. These results provide further insight into support system design, and highlight the importance of considering these factors during development.

- Chapter 6, entitled *'The Brain Dynamics of Contextual Semantic Comprehension'*, presents a study on the detection of the processes of semantic comprehension within the brain, with the results identifying the capability for BCI systems to leverage these biomarkers for integration into an SA support system.

- Chapter 7, entitled *'Conclusions and Future Work'*, concludes the thesis by summarising the key findings and contributions of the contained work, as well as highlighting research directions for future works to expand upon in order to develop a fully functioning closed-loop BCI system for SA support.

CHAPTER 2

Situational Awareness, Physiological Factors, and Improving Situational Awareness with Heads-Up Displays

2.1 The Current Understanding of the Concept of Situational Awareness

Situation Awareness (SA) forms a foundation of the decision making process[7], with an individual's SA forming their comprehension of the situation, and helping to determine what actions they must take in order to manipulate the state of the environment towards their end goal [5]. This importance has led to attempts to augment it [14, 15, 16], as a higher situational awareness improves task performance [17, 18], as well as reducing the severity of errors [19]. There are a number of factors that can negatively affect an individual's SA[20, 21, 22, 4]. To counteract these factors, efforts have been taken to develop techniques and technologies that aim to improve SA.

Whilst the field of modelling SA is firmly established, with models from 1995 [4] still being widely recognized in modern times [23], discourse regarding the models continues to the modern day [24, 25], with the original authors offering clarifications when others in the field provide examples of where misinterpretations may occur [26]. This cycle of refinement has led to a developed field with a number of well-recognized models [27, 24], as well as a collection of less recognized modern models challenging the perceived inconsistencies with the more established models [28, 29, 25]. However, while the discourse regarding the semantics and exact process continue, all of the widely recognized models appear to build upon a base of an Input-Process-Output cyclical pipeline [4, 6], which suggests that an extraction of the agreed upon core principles of SA is possible after analysing the structure and process of the widely recognized models.

2.1.1 Models of Situational Awareness

Endsley's Three-Level Model [4] is often considered the most popular model of individual SA, as well as serving as a base for Team SA, with an understanding that Team SA consists of a collection of individuals with their own individual SA [30]. Endsley's model suggests that SA is a product of a cyclical process that Endsley [26] refers to as situational assessment, and that SA can be broken down into 3 levels, those being 'perception of the elements in the environment', 'comprehension of the current situation', and 'projection of future status' [4]. These levels are worth noting as they show a transformation from information to knowledge as the information about the environment is perceived by the individual, and then the individual proceeds to process and comprehend the information, gaining a knowledge and understanding of the situation. This understanding can then be applied to predict what the future status of the situation will be. It is important to note that whilst Endsley does consider decision making on how to influence the situation as part of the decision making process, it is not included as part of the situation assessment process [7].

Smith & Hancock's Perceptual Cycle Model [6] takes an alternate approach, defining SA as 'adaptive, externally directed consciousness', and suggests that Neisser [31]'s perceptual cycle model can provide a framework for understanding how SA works. From this, we get the cycle that available information in the environment modifies the agent's knowledge of the situation, which then directs the user to action, which samples the environment, potentially changing the available information. It is worth noting a major difference between this model and Endsley's, in that this model considers acting upon the knowledge an integral part of the SA cycle, whilst Endsley places much more emphasis on the perception of the environment and comprehension of information, treating decision making and acting as part of the decision making process rather than the SA process [4]. This would fall under the Environment Modifies Knowledge section of Smith & Hancock's model. This illustrates that potentially one of the major causes of discourse in the SA field is that of scale, and whether SA is just the internal understanding of the situation, or whether the actions an actor may take within a scene is also an integral part of SA, rather than just a result.

Bedny & Meister's Interactive Subsystems Model [5] serves as another alternative view to SA, applying the Russian Theory of Activity to attempt to describe and explain SA in a more comprehensive manner than other models, advancing the motion that "SA phenomena can only be understood as part of the structure of a comprehensive theory of activity" [5]. Under this theory, activity can be divided into three stages: Orientational, Executive, and Evaluative. Due to Orientational activity, individuals "develop a subjective model of reality from which they actively extract distinct representations" and as a result "a dynamic picture of the world is formed". Bedny and Meister [5] also proceed to elaborate that this dynamic picture "provides a meaningful and coherent interpretation of reality and anticipation of future state of the situation". This can be directly contrasted to Endsley [4]'s model, with "extract[ing] distinct representations" forming Level 1, the "meaningful and coherent interpretation of reality" forming Level 2, and the "anticipation of future state" forming Level 3. Furthermore, Executive activity can be directly compared with the "Decision" and "Performance of Action" steps within the model of dynamic decision making that Endsley [4] proposes, featuring "decision making and performance of actions" [5] that are "directed to achieve a conscious goal". Finally, the Evaluative activity can be compared to perception of changes in the environment due to the action (Level 1). This does however raise a potential inconsistency with the Interactive Subsystems model, with Evaluative activity being an "assessment of the result of the activity produced by information feedback" which "leads to corrective actions". It is, however, to achieve the same result of this stage by considering the Evaluative step as another set of orientational and executive activity, with an individual extracting the changes in the "distinct representation" of their reality, altering the "dynamic picture of the world", which then results in a decision to take actions towards achieving the desired goal. When presented like this, the question can be raised as to whether the evaluative step is distinct, or whether it represents another cycle of the entire process with modified understanding and information from the previous cycle.

The models described above are compared below in table 1 in order to compare and contrast the different definitions of SA as well as the identified processing pipelines to form a unified definition of SA as well as a unified theory on the core of the SA pipeline.

Model	Definition of SA	Processing Pipeline
Endsley's Three-Level Model	"The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future"	Perception -> Comprehension -> Projection -> Perception...
Smith & Hancock's Perceptual Cycle Model	"Externally, directed consciousness" that is an "invariant component in an adaptive cycle of knowledge, action and information"	Available information MODIFIES Knowledge WHICH DIRECTS Action WHICH SAMPLES Available information...
Bedny & Meister's Interactive Sub-systems Model	"One important functional mechanism of reflective-orientational activity, which provides a conscious and dynamic orientation in the situation"	Orientation -> Execution -> Evaluation

Table 2.1: Definitions of SA Models

This is vital in order to develop a method of augmenting SA, as it allows an insight into the core factors required for SA, as well as relevant sub-processes that can be aided in order to increase SA.

As can be seen in Table 1, the core processing pipelines for each model are quite similar, with all of them having a three-stage process at the core of the model, with that being perceiving the environment for information relevant to the situation, comprehending this information to form an understanding of the situation, and formulating a plan of and undertaking an action, or series of actions in order to manipulate the situation into a more desirable state. From this similarity and agreement between the models, we can identify the core SA cycle: Perceive ->Comprehend ->Act.

Just as the singular internal model for SA is important for understanding how an individual generates an understanding of the situation around them, so too is modelling how multiple actors within a situation share information and knowledge regarding the situation to ensure the actor's goals are achieved. This information transference allows multiple actors to collectively increase their individual SA, and thus decrease the likelihood of an error from inadequate SA occurring [24, 7].Taking this concept of information transference further, recent models have suggested that non-human agents such as technological artefacts have some level of SA within a scene [32]. This concept is extremely relevant in modern society, where the use of technological aides in complex situations, such as aircraft cockpits, are commonplace. It is important to note that, depending the model you follow, in order for an actor to hold SA it must not only be able to hold information from the environment, it must also have knowledge or understanding of the information [4, 6], as well as in some cases the ability to take action in the environment [6]. That is to say, in order to be considered an actor, one must have both the ability to comprehend the information one holds, as well as the ability to act on said knowledge. To this end, it can be argued that whilst a component is able to display some qualities of situational awareness (Take, for instance, a cockpit airspeed indicator display the airspeed), it would not be until the component hold an understanding of the relevance of its information (If the airspeed is too low, a stall is occurring), and in some cases, able to act (The indicator activates the Stall Warning Light), that a component can be considered an actor in a scene. To this end, it can be argued against Stanton [33]'s statement that 'technological artefacts have some level of SA (at least in the sense that they are holders of contextually relevant information)', as unless the artefact has the capability to understand the information it holds, it is unable to fulfil one of the fundamental steps of most well-accepted SA models.

2.1.2 Adverse Factors in Situational Awareness

There are a number of factors that can negatively affect an individual's SA, such as fatigue [20], high mental workload[21], distractions within the environment[22], and stress[4]. This negative impact on SA is extremely relevant, especially in dangerous situations where a loss of SA could potentially lead to harm or even loss of life. However, to counteract these factors, efforts have been taken to develop techniques and technologies that aim to improve SA.

2.1.3 Factors of SA Loss

From our definition, it can be seen that the process of forming SA has three main components, those being the perception of information, the comprehension of that information to form an understanding of the situation, and a prediction of the consequences that actions taken will have upon the situation. These three components of SA can be compared with the three levels of SA from Endsley's model for the sake of Jones and Endsley [34]'s analysis of errors within both flight crew and air traffic controllers. An alternate classification of errors can be seen in Durso et al. [19]'s assignment of psychological mechanisms to aware and unaware air traffic controllers in regards to SA errors. Durso et al. identify that these psychological mechanisms can be broken down into four categories: Perception, Attention, Memory, and Thinking. These categories focus more on the information acquisition and retention, with Perception, Attention and Memory all fall under Level 1 SA within Jones and Endsley [34]'s classification, while Thinking covers "poor judgement, reasoning, planning, false beliefs, misinterpretations, lack of understanding, or erroneous assumptions" [19], which covers both Level 2 and Level 3 SA from Endsley's Model. This focus on Level 1 SA, combined with the dividing of erroneous operators into "aware" and "unaware", based on whether the operator was "aware that an operational error/deviation was developing" [19]. Durso et al. [19] note that aware operators tended to make thinking errors, whereas unaware operators tended to make more memory and perceptual errors. This, combined with the findings that most of the aware-thinking errors were often classified as Inappropriate Use of Displayed Data, shows a clear importance in ensuring that information is delivered in a manner that is easily parsable, in addition to ensuring the user is aware of the delivery of the information.

From this coverage of the varying types of errors, combined with discussion on the main factors of SA loss by Endsley and Jones [7] and Durso and Gronlund [35], it is possible to focus onto three main areas where SA loss occurs - These being Attention, Memory, and Understanding. Each of these areas will be covered and contrasted with Endsley's and Durso's identification of SA loss factors in order to identify specifics within the areas that should be focused on.

2.1.3.1 Attention

Attention in this refers to the ability to perceive the environment and ensure that perception is not misplaced. Endsley and Jones [7] refers to two major factors of SA loss in regards to attention, those being 'Misplaced Salience' and 'Attentional Tunnelling'.
Misplaced Salience refers to the issue of a user's attention being caught by certain types of information such as flashing lights or bright colours, potentially distracting a user and preventing them from noticing useful information due to their attention being caught by more salient, yet potentially useless, information.
Attentional Tunnelling serves as another other core factor of SA loss, where a focus on a specific aspect of the environment causes an individual to become unaware of changes in their surroundings due to the 'tunnelling' of their attention, resulting in a loss of

awareness in the unattended areas. This factor can lead to identified phenomena that display a loss of SA, such as Inattentional Blindness. Whilst Endsley and Jones [7] does not refer to this, it should be noted that there is a lack of a reference later than 1997 in the relevant chapter. It was not until a year later that Mack and Rock [36] coined the term 'Inattentional Blindness'. whilst the factor itself has been referred to as 'Attentional Narrowing/Tunnelling' since 1943 [37, 38, 39].

2.1.3.2 Memory

Memory in this form refers to the ability to store and recall information relevant to the situation. This ability however is extremely limited. It was originally theorised that roughly seven, plus or minus two chunks (related pieces) of information were able to be held in working memory at once [40], however this number was later considered to be inflated due to the use of alphanumeric stimuli, with a more accurate number being approximately two to four for visual or auditory stimuli, with memorisation being even more difficult when the stimuli are novel[41, 42]. This highlights a complex interaction when considering memory as a limiting factor, as how well learnt the stimuli are, as well as the method in which they are delivered, can greatly alter the ability to retain them in memory. In addition, this information is not permanent, as without active work to keep the information in memory, the information will quickly become lost [43]. Therefore, reliance on short-term memory can lead to errors should the individual be forced to remember more information than they are capable of, or should they be forced to retain this information for a long period of time. Endsley and Jones [7] describes this as a 'requisite memory trap' when systems or situations rely heavily on an individual's short-term memory for successful function. This loss of information is significant as it can be a root cause of higher-level factors regarding comprehending the situation such as forgetting information that was used to correct an errant mental model. This is potentially an oversight by Endsley and Jones [7], as whilst these factors appear distinct, they can be the causes and consequences of other factors, with those at the perception and comprehension levels having a cascading impact onto later stages.

2.1.3.3 Comprehension

Comprehension can be broken into two separate areas, these being the parsing of the information into an understanding of the information, and the formulation of a set of actions that will manipulate the situation into one more favourable for a desired outcome. Whilst in a unified definition these fall under both Knowledge and Action (Level 2 & 3 of

Endsley's Model), the factors affecting both remain much the same.

In regards to the transformation of perceived information to knowledge, there are two main factors that can cause a degradation in SA, these being Data Overload and Errant Mental Models. Data overload refers to how a person can only take in a limited amount of information within a certain time-span, with an individual either having to filter the information for the most relevant pieces (which itself costs mental resources), or take a longer time-span to process this data. In this situation, it is possible that the individual may miss relevant information, or fail to process important information within a time-limit, resulting in a loss of SA, especially in time-critical situations. The second factor for information processing is that of Errant Mental Models, wherein information may be misinterpreted due to an individual not understanding the meaning of the information they have received. This is a severe risk to SA as not only can the knowledge formed be incorrect, but the individual can also be under the assumption that their understanding is correct and any conflicting information is erroneous.

For Action, there can be two factors identified within Endsley and Jones [7]'s discussion, these being Complexity Creep and Out-of-the-loop Syndrome. Complexity Creep can affect the parsing of information as well as formulation of action, and can also be a compounding factor with others, as a complex system requires a more sophisticated mental model, thus increasing the likelihood of an errant mental model being formed. The other factor, Out-of-the-loop Syndrome occurs as a consequence of automation, where an individual's understanding of the actions taken by an automated system differs from the actual actions taken, leading to the individual holding an incorrect understanding of the situation. Furthermore, the individual may neglect to monitor the automated system, instead devoting their attention to non-automated tasks. This can itself lead to a form of inattentional blindness, where the individual is blind to the actions the automated system is taking. This can lead to issues in the event that the automated system reaches a state that requires user intervention, as the user may be unable to detect the problem or interpret the current situation due to lack of information regarding the automated task.

2.1.3.4 Stressors

Stressors (Referred to as WAFOS - Workload, Anxiety, Fatigue and Other Stressors - by Endsley and Jones [7]) refers to factors, both psychological and physical, that can place a strain on SA through a multitude of compounding factors, exacerbating the previous identified issues to cause an even greater loss of SA. Relating to perception, stressors reduce an individual's ability to gather information effectively, as well as reducing attention paid towards the peripherals of the senses, leading to both a higher risk of missing information, as well as an increased chance to succumb to inattentional blindness.

In regards to comprehension, stressors can be dangerous, and often distracting, leading to an individual's likelihood to remember its existence, taking up the already limited space in working memory, furthermore it has a detrimental impact on both the formation and retrieval of information from short-term memory [44].

Finally, in regards to action, stressors can cause an individual to arrive at a conclusion and act without taking all the available information and knowledge into account, in an action termed premature closure. This rapid and uninformed decision making can lead not only to decisions that are sub-par for achieving an individual's goals, but can actively harm their chances as the action may have negative consequences that the individual may have become aware of had they sufficiently analysed their proposed actions.

2.1.4 Examples of SA Loss

There exist a multitude of phenomena that represent a loss of SA, such as Inattentional Blindness, Change Blindness, Inattentional Agnosia and Inattentional Amnesia. It should be noted that the described phenomena do not all have the critical point of failure within the perception stage, nor do they all necessarily only have one root cause. For example, Inattentional Amnesia could be caused by both Attentional Tunnelling resulting in the information being skimmed over, and thus the Requisite Memory Trap discards this half-parsed information as irrelevant. Another key example would be that Change Blindness does not necessarily have to be caused by a lapse in short-term memory, but could potentially be due to the pre-change object having not been thoroughly attended to, and only the post-change being observed, as identified by Hollingworth et al [45, 46].

2.2 The Use of Augmented Reality & Heads-Up Displays for Situational Awareness Improvement

Augmented reality refers to "the use of computer displays to add virtual information to a user's sensory perceptions"[47]. By providing an individual with virtual information, a designer is able to guide attention towards relevant information by capturing attention through the use of salient events during data presentation [48].

Failure Type / Stage	Perception	Memory	Comprehension	Projection
Inattentional Blindness	X			
Inattentional Amnesia	Y	X		
Inattentional Agnosia	Y	Y	X	
Change Blindness	Y*	X	X	

Table 2.2: Identified Stages vs Failures of Awareness | Y = Stage Completed Successfully, X = Failure in Stage, Blank = Stage not reached

The use of Augmented Reality (AR) as an aide is quite common, both a method of delivering information that an individual may have missed or was not able to perceive [49], as well as allowing them to receive information without having to take their attention away from the environment [50], with many specialized display systems existing for information delivery, each designed for their own specific use-case [16, 1, 51, 52, 53, 54, 55, 56, 57, 58]. However, these systems often require specific training in order to understand how to use them, as quite often these systems are dissimilar to commonly implemented information delivery systems, which can prove to be an issue if enough training is not done on the system [7].

2.2.0.1 User Interface Design

How the information is presented to an individual is a core aspect of any guidance system, as poorly presented information can increase mental workload, leading to a higher rate of error [7]. Furthermore, where the information is presented in the field of view alters how the information should be displayed. For instance, peripheral vision is not suited to detailed information, but motion in the periphery is usually noticed by an individual [59]. Additionally, it is often useful that if physical motion is occurring, the display itself mimicks the motion, to aid the user in forming an appropriate mental model [60]. When designing for AR, additional considerations have to be taken for issues that are unique to Mixed/Augmented Reality [61].

Additional parallels can be drawn between designing for stressful environments and designing for video games. Endsley [7] discusses how display complexity can make it more difficult to perceive needed information by increasing the time needed to search for said information. This can be seen mirrored in game heuristics, where a listed heuristic is to "Provide visual representations that are easy to interpret and that minimize the need for micromanagement" [62], in order to minimize the time needed to search for needed information.

2.2.1 Improving SA through Visual Aides

Previous studies have investigated various methods at providing information through AR to aid in target localization, whether that be by designing new display techniques such as EyeSee360 [63] or by adapting and investigating the use of display methods from other domains [64]. However, it has been shown that higher visual complexity can lead to a worse performance in target localization [65], which has led to investigation as to whether other methods such as audio-tactile guidance could be used, which, whilst increasing the likelihood of awareness of a stimulus, increased the time taken to accurately locate the stimulus [66]. This leads to the continued use of visual guidance when time-to-locate is a relevant factor in system design. Previous research has also been done specifically on the use of radar-like [67, 68] and compass-based [69] designs to support navigation and localization, although to the best of the authors' knowledge no previous work has compared these HUD designs against one another whilst also considering video game experience as an indicator towards usability.

2.3 The Current Understanding of the Effects of Stress on Situational Awareness and HUD Use

2.3.1 Effects of Stress on Situational Awareness

Stress has been shown to have a number of detrimental effects on human performance, ranging from memory function [70] to visual processing capability [71], with stress and other similar factors being noted as a danger to operator safety [20, 37].

This criticality in ensuring operator safety has led to a wide array of works in measuring and augmenting human performance [72, 73, 7], with one of said methods of augmentation being providing information through the form of a HUD or some other form of Augmented Reality ranging from worn or held systems[54, 1] to those integrated into the machinery or vehicle that is operated [74, 75]. However, whilst it has been known since the 1980s that such methods of augmenting performance can, in some situations, be detrimental [76, 9], and efforts have been made to identify the functional requirements of HUD systems[77], the behavioural effects of stress on modern user interface design for HUD effectiveness has yet to be investigated.

2.4 The Current Understanding of the Effects of Circadian Fatigue on Situational Awareness and HUD Use

There are many factors that can affect one's situational awareness, with fatigue being a major factor in human performance and awareness of their environment. As certain professions require long hours over the course of multiple days, such as oil rig workers or firefighters, fatigue can become a large factor in these long-term deployments in the workplace, and therefore mitigating the effects of fatigue becomes critical in ensuring adequate performance in task completion. While studies have been done on the effects of fatigue in regards to computer use and task completion, very few have investigated the effects of fatigue on AR assistance for audio localization. As perception is a core component of SA, it is critical that individuals receive the support they need for task completion. However, as more information is provided in an audio-visual form, it must be ensured that such a support system does not overload the user or cause performance degradation.

As AR has been identified as a method of providing a visual aspect to auditory aspects of the environment, HUDs have continued to be used to provide localization assistance. However, the long-term effects of fatigue on HUD usage has been primarily limited to stationary use, such as that in trucks or aircraft, where the user is not required to move within their environment, and is solely focused on the task at hand, normally the piloting of the vehicle. Therefore there is an identified need to investigate the effects of fatigue on personal HUDs in a realistic environment, where awareness of your overall surroundings may not be the primary task.

In order to investigate these effects of physiological factors on SA performance, the effects of circadian fatigue (sleep deprivation) was investigated. Additionally, the previously used AR support system was also investigated for suitability in improve performance when in a state of circadian fatigue.

2.4.1 Effects of Fatigue on Human Performance

Circadian fatigue, or sleep deprivation, has been shown to have a significant impact on human performance, from physical tasks such as driving [78, 79] to cognitive tasks such as multi-tasking [80] or standardized tests [81], with this change in performance capability even causing changes at the neurological level [82, 83]. This link between cognitive capability and fatigue state has been identified as a metric of gauging fatigue level, with the possibility of predicting fatigue state being represented in literature [84, 85], with a number of different different target features being identified as possibilities [86, 87, 88, 89]. However, these systems designed to measure physiological states are often large, with a long setup time, limiting their capabilities to stationary simulators, or a reduction in data quality for portable systems.

To protect against the negative effects of circadian fatigue, there have been a number of studies and reports into the effective safety practices in dangerous environments [90, 91], as well as regulations put in place in order to ensure safe operation of dangerous equipment [92]. In addition, more portable systems have been developed that can classify a user's fatigue state, such as systems that can make use of a minimum of just sleep data [93], reducing the biometric data and equipment required to gauge a user's fatigue state.

2.4.2 Effects of Fatigue on HUD Effectiveness

Heads-Up-Displays (HUDs) are one method of providing additional guidance to individuals to aid in task completion, and whilst non-visual guidance methods have been investigated [66], the traditional paradigms still make use of visual elements. In addition, when these HUDs are integrated into a head-mounted display, additional factors such as limited field-of-view must be considered [64].

While HUDs have been implemented in many systems, there has been a long-time understanding that HUDs can cause their own cognitive issues [76, 94]. Furthermore, research into appropriate techniques for situations where fatigue may be a factor, such as flight decks, has been investigated [95, 9]. While this research in identifying the functional requirements of HUDs has continued to the present day [77], with wearable HUDs becoming more available for widespread use, an investigation into the effectiveness of wearable HUD systems must be investigated, even more-so when used in situations such as field maintenance of equipment [96], where fatigue from long hours can be an extremely relevant factor.

2.5 The Current Understanding of the Process of Comprehension within the Brain

2.5.1 Comprehension in Psychology

Farhadi et al. [97] identified that in regards to object recognition an object can be defined by its' component attributes, however their study was limited to the use of "sensory attributes" [98] and only made use of the physical features, and did not make use of functional or encyclopedic attributes. This leads to a surface level comprehension of an object, but higher-level comprehension is lost, such as being able to state that a pigeon *has wings*, but not being able to state that it *can fly* or that, by flying, it *is safe from (most) ground based predators*. Additionally, even while this simple logic could be implemented in a system for higher-level comprehension by a computer vision system, these rules are not always true, depending on context. For instance, an ostrich has wings, but cannot fly, and whilst a chicken is widely accepted to be able to fly, it does not fly in the traditional sense, leaving it vulnerable to ground-based predators. Whilst it is possible to account for these semantic edge cases, both the complexity of the relationships as well as the potential for an extraordinarily large number of relationships and potential attributes leads to difficulties in representing all the potential attributes an object could be assigned.

However, work has been done on identifying these more abstract semantic properties in the field of cognitive neuroscience through identifying when an individual mentally assigns an attribute to an item.

2.5.2 The Brain Dynamics of Object Recognition, Identification, and Comprehension

In a study by Chan et al. [99] which made use of EEG and MEG, participants were provided with a stimulus word, represented in either an auditory or visual format, and were required to press a button if the object represented by the word was larger than one foot in any dimension (e.g. sofa) whilst not pressing the button if the object was less than one foot (e.g. cricket). During analysis a classifier was trained to not only classify based on the large/small distinction, but also on a living/non-living distinction, as well as an individual-item distinction. Results showed that the classification of living/non-living as well as individual words were both significantly above the chance accuracy threshold. Whilst these results are promising semantic comprehension, arguments have been made that the attribute of animacy is considered 'special' in that incongruencies in regards to animacy elicit a higher amplitude response under EEG than other semantic incongruencies [100], as well as plays a larger role in processing of object relative sentences overall [101]. Additionally, recent studies have also shown that the animacy features of nouns may be predicted ahead of noun presentation during language comprehension [102]. This means that whilst animacy may prove to be a detectable semantic attribute through the use of EEG, it may not be truly representative of high-level, abstract, or opinion-related semantic attributes.

A counter-argument to this can be presented in that the processes for lexical semantic comprehension are potentially different from those that occur when presented with the actual object instead of a lexical representation of the object. However, a study similar to the previous one, conducted by Carlson et al. [103], made use of images instead of lexical representations, as well as consisted of a main task irrelevant to the semantic classification, in such a way that the object images were only displayed as the background to the main task. The results showed that once again a meaningful classification accuracy could be achieved in the category of animacy. Furthermore, the classification of inanimate objects into naturalistic or man-made was also possible, which further points to the possibility of higher-level semantic categories and attributes, in this case natural vs manufactured. This is further supported by a study that investigated the semantic categorization of unknown sounds into either living or man-made objects, which found that "the brain is able to discriminate the semantic category of environmental sounds even when this does not transpire behaviourally" [104]. Additionally, as shown by Correia et al. [105], bi-lingual lexical representations of animals could be discriminated, however "significant across-language generalization was possible around 550,Äì600 ms" potentially distinguishing the semantic representation of the word, apart from the lexical representation itself. Finally, the study by Carlson et al. [103] can lead to the conclusion that semantic comprehension of an object can occur unprompted due to the experimental paradigm, thus giving the potential for the use of a biomarker in a passive BCI, should the biomarker for higher-level semantic classification continue to occur even when unprompted.

2.5.2.1 Brain Regions

In regards to brain regions where semantic activity takes place, semantic classification and categorization processes have been shown to have late positive components in the occipital and left temporal regions of the brain. This ties into semantic memory processes, as in order to categorize an item, one must first retrieve their understanding of an item from their memory. This can be seen through fMRI studies which show the flow of information when perceiving or reconstructing an image of an item from memory [106, 107] or when categorizing animacy of a stimulus [108].

Additionally, an investigation into complex natural speech by Huth et al. [109] has revealed that comprehension of lexical semantics is distributed throughout various regions of the brain. To investigate this they had subjects listen to multiple hours of narrative stories, whilst using fMRI to measure the brain activity. The words in the stories were assigned lexical categories when they represented an item that fit within one of the categories (The categories being "tactile (a cluster containing words such as "fingers"), visual (words such as "yellow"), numeric ("four"), locational ("stadium"), abstract ("natural"), temporal ("minute"), professional ("meetings"), violent ("lethal"), communal ("schools"), mental ("asleep"), emotional ("despised"), and social ("child")" [109]). Whilst some of these categories, such as visual or numeric, consist of more concrete information in that the attribute within the category will not change based on the context, categories such as violent or emotional have attributes that may or may not be assigned to an item depending on the context, such as a person believing an object to be safe when it is actually dangerous.

A further study by Murphy et al. [110] found that when presenting participants with grey-scale images of mammals and tools, and asking participants to silently name the shown item, they were able to categorize the participant's EEG signals into which category the viewed stimulus fell, with significant accuracy, through the use of a Support-Vector Machine. Whilst they discovered that the optimal classification timeframe and frequency for classification was 100-370ms and 3-17Hz respectively, most interestingly they found that "semantic processing and the resulting representations are widely spread across the brain, in a fashion that is somewhat shared between the participants that took part in the study". This reinforces the findings of the study by Huth et al. [109] that semantic representations are distributed throughout the brain. However, this shares one major similarity with Huth et al. [109]'s study that distinguishes it from other semantic categorization studies, in that participants were asked to silently name the object. whereas in other studies participants were simply asked to indicate whether an object fell into a category or not, without explicitly naming the object. This addition of lexical processing, both in listening to audio in the case of Huth et al. [109] and the silent naming in the case of Murphy et al. [110] could potentially be related to the activity in the frontal cortex, which is often associated with lexical processing. This possibility is further supported by Chan et al. [99]'s study, where the auditory version of a stimulus had a much larger and longer activation within the frontal cortex when compared to the visual representation of the word. This longer activation, as well as change in location, is supported by Sperber et al. [111] whose study identified that visual to visual priming had a much more significant effect than lexical priming, which led them to hypothesise that in visual priming physical features of the object were used to aid the semantic comprehension, which provides support for the shorter activation and larger activations in the occipital and temporal regions in the case of visual stimuli [99]. Finally, a review of 120 functional neuroimaging studies identified that the reliable areas of activation across these studies found that these areas formed a "distinct, left-lateralized network" in regards to the semantic system, though also note that the relevance of the prefrontal cortex, as well as identifying some activity in the right hemisphere related to semantic comprehension [112].

2.5.2.2 Timeframes

As semantic comprehension has been shown to be related to the recollection of relevant information from semantic memory, potential insights into the timeframe in which semantic comprehension takes place can be gained by investigating the previous research into recollection from semantic memory.

A study by Costanzo et al. [108] identified that when presented with both an audio or visual stimulus representing a living or non-living object, a late positive component was evoked in the parietal region of the brain, ranging from 400-800ms for visual images and 550-900 for audible words. Whilst the amplitude of this ERP was not significantly different, there was a significant difference in latency between the two categories of object, with the maximal response significantly earlier for living items as opposed to inanimate items. Whilst Costanzo et al. [108] did not classify their results, the previously discussed study by [110] classified their results found that the optimal classification window for mammals or tools was 100-370ms, identifying the ability to classify objects based on the early ERP components as well as the later components, in fact finding that classifying using the earlier components led to a higher classification accuracy. Whilst arguments can be made that "tools and mammals" are different categories than "non-animate and animate", the attribute of animacy has been shown to have an important role in semantic processing [100], with neurological processes using the attribute of animacy to aid complex lexical processing [101], and even predicting the animacy of upcoming nouns [102] in language comprehension. This does unfortunately leave an ambiguity in [110]'s study as to whether the classification was for their specifically listed categories or for the attribute of animacy as no "non-tool non-animate" or "non-mammal but animate" stimuli were used. This is however mentioned, with it being mentioned that "semantic categories used in studies of this kind are often somewhat arbitrary and so can present problems for the interpretation of results" [110]. In more recent studies, efforts have been taken to mitigate these problems in interpretation through ensuring that presented stimuli have a 'nested' structure, where multiple categories can be assigned to one stimulus, such as in a study by Bainbridge, Hall, and Baker [107] where there are tools and non-tools, that can be categorized into big and small, which represent objects. This nested structure where each branch contains the same sub-categories helps to alleviate (though not prevent entirely) potential problems from occurring.

2.5.2.3 Identifiable Components

Continuing to draw on the link between semantic memory and semantic comprehension, a study by Klimesch, Schimke, and Schwaiger [113] combined two tasks to identify the key differences between semantic and episodic memory recall. In order to do this, for semantic memory participants were asked to identify whether a concept-feature pair (e.g. "lark-sings") was congruent, then afterwards, for episodic memory, participants were required to identify whether a presented concept-feature pair had been presented before. Klimesch, Schimke, and Schwaiger [113] found a distinct difference between the two tasks, with semantic memory activity taking place primarily in the upper alpha band. Unfortunately however, this study focuses on the congruence of pre-provided pairs, as opposed to participants providing the pairing themselves. This, as well as how the data was not classified by "superordinate categories" such as bird or weapon, even though stimuli were grouped as such, limits the relevance of this study to identifying the band in which semantic recollection primarily takes place. Some of these questions were addressed in a follow-up study by Klimesch et al. [114] which added an additional free-association task, as well as repeating a task similar to the semantic memory task from the previous study. The follow-up study found that judgement of semantic congruency in pre-provided pairs had "pronounced increase in desynchronization which is much stronger over the left hemisphere for frontal, central, temporal, and parietal recording sites" in the upper alpha band, whereas the free-association task had a "right hemispheric advantage" that was "most pronounced at parietal recording sites". This study both confirmed the previous finding that the upper alpha band responded selectively to the demands of the semantic memory task. The study also found that whilst the occipital region responded to both semantic and visual processing demands, the other discussed regions of the brain responded only to semantic processing demands.This further supports the concept of distributed semantic processing, with more complex or free-form tasks making use of the right-hemisphere whilst simple congruency tasks making use primarily of the left-hemisphere.

It must, however, be noted that Klimesch et al. investigated only thematic semantic relationships in their study (e.g. a lark is thematically related to singing), whilst other types of semantic relationships, such as taxonomic where there is a shared category membership, exist. Klimesch et al. do actually identify this membership through their use of superordinate categories to group stimuli, though these categories are not investigated, and instead used to prime participants to prevent misunderstandings, such as presenting the category of "birds" to prevent ambiguity as to whether "swallow" is the bird, or the action [113]. To investigate any potential differences between these two categories, [115] conducted a study that investigated the differences in EEG theta and alpha responses in participants when they were presented with word pairs in an audio-visual format that were either taxonomically or thematically related, or unrelated. Participants were instructed to not respond in any way, as previous work had shown that the "only ERP study to elicit differences between thematic and taxonomic relationships was passive" [115] (referring to [116]). Although the study by Maguire et al. found no distinctions in ERP, they did find ERSP differences, with an increase in theta power over right frontal areas for thematic versus taxonomic relationships, and an increase in alpha power over parietal areas for taxonomic versus thematic relationships. Whilst the main finding of the study is support of the idea of a distributed semantic map, as per Huth et al. [109], another interesting finding is how distinction of the type of semantic relationship is limited to passive classification, though both active [99] and passive [103] classifications are able to identify the item being classified, as well as the category within which it is being classified.

2.5.3 BCI Use for Improving SA

Brain Computer Interfaces (BCIs) provide a way for individuals to interact with and control computers and other machines through their brain activity. This can either be in an active form, where individuals actively issues commands and instructions to the BCI, or a passive form [117], where the BCI passively monitors and interprets the user's mental activity without active input from the user. A passive BCI can then, when paired with some form of AI or rule-set, use these interpretations and observations to take the appropriate action when it observes a specific mental activity or the user enter a specific mental state, or to save time by removing the human motor element and pre-emptively taking the action in time-sensitive situations, such as activating a brake in an emergency-braking situation [118].

2.5.4 Physiological Data for BCIs

There exist many methods of measuring the neurological signals from the brain, ranging from invasive procedures where the sensors measuring the brain signals are placed under the scalp to increase the resolution and clarity of the data, to non-invasive procedures where the sensors are simply placed against the scalp, either with a conductive substance between the sensor and scalp to help amplify the signal, or with no conductive substance at all. This poses a large barrier to BCI adoption, as invasive methods are often off-putting and require surgery in order to ensure the sensors are in place, limiting the adoption of invasive methods to the medical domain. This means that for wide-spread and easy adoption such a system must be non-invasive, which poses the issue of resolution and noisiness of the signal.

Another major issue in measuring the neurological signals is that of equipment size in regards to resolution. It is possible to collect high-resolution imaging of the brain, including deep inner regions that other methods may not be able to collect, through methods such as functional magnetic resonance imaging (fMRI) or magnetoencephalography (MEG), however using these methods comes at the cost of such equipment being both expensive and physically bulky, requiring a static setup in order to collect data. Whilst these setups are feasible for high-stress situations in which an individual is tele-operating a drone or robot, it is much more likely that such a system requiring rapid-response SA augmentation would be useful to the many situations in which the individual is physically located in the situation instead. This further limits the available techniques for measuring brain activity, the two most notable being Electroencephalography (EEG) and functional Near-Infrared Spectroscopy (fNIRS).

2.5.4.1 Electroencephalography (EEG)

Electroencephalography is the practice of detecting electrical signals within the brain through the use of an array of electrodes resting against the scalp. EEG is commonly used in BCIs, both due to its relatively high temporal resolution when compared to other acquisition methods [119], as well as the low cost and portability [119]. However, a major issue with EEG is its susceptibility to noise, as well as the lower classification accuracy when compared to stationary methods such as MEG and fMRI, as well as to a lesser extent fNIRS. The field of EEG research is already well established, however, which allows new research to have a large block of existing literature to build upon. Additionally, recent work has aimed to counteract the issue of noise and artefacts contaminating the data through the development of novel and efficient approaches for the effective removal of noise and artefacts from the data [120, 121].

2.5.4.2 Functional Near-Infrared Spectroscopy (fNIRS)

fNIRS serves as another method of detecting brain activity, through identifying the differences between oxygenated and deoxygenated haemoglobin in the brain through the difference in reflected wavelengths in near-infrared radiation [122]. Since it is presumed that an increase in oxygenated blood within the brain is tied to that region being active, this difference can be used to map brain activity over time. However, fNIRS is limited to near-surface regions of the brain, as it has the inability to visualize the subcortical structures of the brain. It is, however, free from movement artifacts that are commonly picked up by EEG electrodes, which many consider a tangible benefit to using fNIRS.

While the discussion about whether EEG or fNIRS is superior in regards to recording neurological data, numerous studies have made efforts to create a hybrid-BCI through the combination of both EEG and fNIRS data. However, whilst this combined system does lead to a higher classification accuracy than either of its constituent parts alone [123, 124, 125, 126], it is both more computationally expensive [127] as well as having a longer delay due to the long time delay of the haemodynamic response in regards to fNIRS [123]. This creates an interesting dilemma of whether to prioritise classification accuracy or time to classify in real-time systems, as well as whether the additional cost of the fNIRS equipment and time delay added is offset the extra classification accuracy, as well as additional robustness in noisy environments [83].

2.5.5 Integrating BCIs with AR SA-support systems

The use of BCIs with relation to SA has grown greatly in the recent years, both due to the recognition of the importance of SA in stressful tasks [12], as well as the growing complexity of said tasks over time [13]. As interruptive tasks such as SAGAT cannot be done in real-time, nor are they feasible in a real-world environment, another way of detecting, and potentially correcting, SA must be achieved in order to provide real-time assistance. A recent study by [3] revealed that when a passive BCI was integrated into cognitive modelling of aircraft pilot's responses to alerts and messages, the classification accuracy rose to 87% from 72% with the model that did not consider the EEG data. This highlights the importance of integrating neurological data into current SA support systems, as a higher success rate in classification of response, as well as the ability to rapidly respond, can help prevent accidents from occurring.

2.5.5.1 Perception

The perceptions of items and aspects of the environment can be determined through neurological data, both in regards to the presence of an object through the detection of P300 and activity in the visual cortex, as well as determining which object is being perceived through the use of a classifier [128]. It can, however, be debated as to whether object recognition falls under perception or comprehension. For instance, say you are driving and you notice a car in the other lane - without adding additional complexity, we can ask whether it was a car was perceived, or whether what was perceived was a large chunk of metal, that was then comprehended as a car. This first reference to semantics, or the meaning of something, will become relevant, not just in simple object recognition, but in regards to comprehending the meaning behind an object and the situation as a whole. Whilst Endsley and Jones [7] states that this comprehension of what each aspect of the environment in isolation falls under perception and that consideration of these objects and the perceived information as a whole comes later as a second step, through neurological investigation into semantics we can see this may not be the case.

2.5.5.2 Memory

Roy et al. [87] found in their study that it is possible to classify both the fatigue state of a participant as well as their mental workload over a large time-on-task. This is important as it shows that we are able to determine an individual's workload, and adjust the information presented, as an increase in mental workload has been shown to lead to a decrease in accuracy in recalling information as well as an increased time to recall [129]. However, whilst there is research into the influence of working memory on the performance of BCIs, Sprague, McBee, and Sellers [130] found that after including "psychological covariates (i.e., hunger, caffeine, fatigue, mood, and motivation)" the influence of working memory capability on BCI accuracy was non-significant. This is important to note as it means as well as attempting to lessen stressors on an individual, lowering the required working memory capacity should also increase the effectiveness of

BCI classification. Interestingly, LaRocque et al. [131] identified that it was possible to decode the information held in short-term memory, which opens up further possibilities of identifying exactly what an individual remembers, and more importantly, what they did not remember or they forgot.

2.5.5.3 Comprehension

Identifying high-level comprehension is rarely considered from a neurological perspective [132], quite possibly due to the complex nature of high-level understanding. In regards to augmenting comprehension for SA, it possible to classify an individual's understanding of an object, including but not limited to what the object they are perceiving is [133] (Or what it represents in the case of words [99]), whether an object falls into a specific semantic category [134] (such as if the object has social properties [109]), as well as if the object is related to a previous object [135]. As comprehension, as discussed by Endsley, is the integration of the data gained from perceiving the environment and deriving meaning from it, we must first identify what 'meaning' means, in a neurological sense. Due to this, a larger section on the comprehension of meaning is provided later, detailing the current state of detecting the neurological biomarkers for comprehension. A major part to note is that whilst many BCIs aim to monitor users for a change in state such as a loss of attention or drowsiness, there remains to be one that detects and acts upon an individual's semantic comprehension of an object or situation, though the comprehension of semantics is steadily being recognized as extremely relevant in the performance of BCIs [136, 137].

2.5.5.4 Projection

Identifying the predictions of an individual is a complex task, made especially more-so by the complexity of the prediction itself. Whilst the identification of surprise, when a prediction is proven incorrect [138] or an unexpected event occurs [139, 140], can be determined, identifying the complex changes that an environment, especially a stressful one, goes through is often much harder. However, it has been shown that P300 ERPs are evoked in prediction tasks for potential stimuli [141], in much the same way that they are evoked for presented stimuli. This leads to a potential question as to whether predicted semantic categories are assigned to an object, when future state is explicitly considered. If so, the relevance of semantic comprehension is once again raised, as not only is one considering what an object currently means, but what an object will mean in the future.

2.5.5.5 Action

The ability to predict movement before it occurs is a potential route to identify an action before it is undertaken [142], however it is often the last available opportunity to interject before such movement is undertaken. Therefore, ideally, one would be able to detect the decision that led to the movement being taken, and interject at the earliest possible chance, in that case the decision being reached. Unfortunately however, only recently have studies been able to predict an accept/decline decision based on single-trial data [143]. However, the use of semantic information relating to an object have been hypothesised to aid classification accuracy [137], therefore it is possible that if it is possible to determine the semantic categories of the action the individual is intending to take, then that may potentially make classification of the actual action more likely due to a reduction in possibilities.

2.5.5.6 Feedback Insight

As noted in Endsley's Dynamic Decision Making model, feedback from the performance of actions alters the state of the environment. This type of action-related feedback can often be used to determine if the undertaken actions were effective in working towards a goal. However, in order to perceive these changes in the environment, one must first understand where to direct their attention in order to notice this information. This once again has relevance to semantics, in that a relationship of meaning between the action and the object or location that will change is established. As previous studies have shown that thematic semantic relationships can be identified, the detection of an erroneous relationship, or the lack of any relationship, could allow a BCI to direct attention towards the areas that may provide additional information for an individual. This further ties into the detection of internal visualization and imagination of objects where the imagination of an object or category [107] could be used to determine what objects an individual thinks will become relevant later on.

2.5.6 Classifying the Data

Numerous classification methods exist in regards to neurological data and BCIs [144, 145, 146], ranging from traditional methods such as Support Vector Machines or Naive Bayes classifiers, to various types of Neural Networks, ranging from Fuzzy Neural Networks to Convolutional Neural Networks (CNNs) to Long Short-Term Memory Networks (LSTMs), or any combination of the above, such as the combination of CNNs and LSTMs to form cLSTMs. Further common methods include both Linear Discrimination Analysis and Canonical Correlation Analysis. However, as the number of different classification methods are so large, an attempt to discuss them, especially in regards to real-time classification, is beyond the scope of this review.

2.5.6.1 The Use of Semantic Comprehension in BCIs

Detecting and classifying both semantic comprehension and classification is of great use in the field of BCIs. As briefly discussed by Simanova et al. [147], the categories of imagined or perceived objects could be used to actively control or communicate with a BCI. More recently, however, there has been investigation into whether the categories from the study by Huth et al. [109] could be detected using EEG instead of fMRI [148], for use in Silent Speech BCIs [137, 149]. The concept proposed by Rekrut et al. [137] involves using a preliminary classification of the internal monologue in order to determine the semantic category of a word prior to word classification, to aid both classification accuracy, as well as the number of words that a Silent Speech BCI can detect. However, as is shown below in Table 2.3, whilst previous works have achieved a meaningful classification accuracy, it is unfortunately not high enough for implementation in real-world applications. This leads to a noticeable gap in the field, as whilst 40% accuracy demonstrates a significant ability to classify over the expected 20% random chance, for real-world application in potentially dangerous situations, the accuracy must be significantly higher to see a chance at real-world adoption. To further add to the difficulty of classification, semantic processing within the brain has been found to be distributed throughout the brain, as shown by a meta-review of 120 papers covering semantic processing [112]. This widespread distribution can be seen as a large barrier to classification, especially using a wearable system, both due to the number of sensors required to effectively measure the numerous brain regions, as well as maintaining the quality of the collected data. This highlights the requirement of locating a smaller region of the brain to focus on for data collection, in order to ensure that a man-portable system would be able to function.

Another potential use-case for semantic classification that remains to be explored is the use of BCIs to detect and correct errors in semantic comprehension. As discussed previously, high levels of stress and workload can lead to a reduction in situational awareness, which can lead to errors with potentially disastrous results. As of yet though, we lack the ability to identify an individual's understanding of an object or a situation, with the closest current capability being able to determine whether an individual is confused [150]. This means that a situation in which an individual forgets to consider a category, such as an electrician forgetting to consider whether the mains power was on or off, or even worse, mis-comprehends it, such as it being safe due to the power being off, when it is in fact dangerous as the power is on.

Without the ability to identify a user's comprehension of the situation or their projection of the current state, such a system is restricted to the Action stage of SA when the model can begin to predict a user's action from the movements they have begun to take. This large delay, and abrupt interruption when the individual has already begun to undertake a course of action could prove to be detrimental, as they will have to reassess the situation after correcting their errant mental model [7], which in a stressful situation, they may not be able to sufficiently do if time constraints exist. This means that the identification and correction of errors at the earliest possible opportunity is of utmost importance in time-constrained situations to ensure both the success of whatever action is being undertaken, as well as the safety of any individuals involved. The integration of semantic comprehension into passive BCIs such as Direct Sense BCIs [119] adds an additional layer of integration between human and computer. For instance, in the "BCIs meet wearable computers" section, Lin and Do [119] discuss the use of augmented reality headsets, namely the HoloLens 2. Potential integration could be done between an AR headset and a BCI to allow, through the use of external cameras and gaze tracking, the headset or a paired processing unit to identify what the user is looking at, and with the BCI determine their understanding of the object and its underlying attributes. The AR display could then be used to provide information to either correct erroneous understanding or alert the user to anything they may have missed.

2.6 Identified Gaps in the Field

From the above literature review, it is possible to identify two major gaps in the literature regarding Situational Awareness. Firstly, there is a noticeable gap in research regarding the effects of physiological factors on HUD effectiveness, specifically in regards to head-mounted systems, as the primary focus of prior research is on aviation and ground vehicle systems [74, 75]. In order to ensure the effectiveness of worn systems under physiological strain, the efficacy of these systems must be tested, which is a gap that can be filled by further research.

Secondly, while a significant amount of research has been done into the neuroscience behind perception, research into the brain dynamics of comprehension has yet to be studied thoroughly [132]. This lack of understanding poses a significant problem to development of a comprehensive system, and whilst the research required to ensure an effective system is vast, ranging from identification of a biomarker to ensuring accurate classification, efforts can be made to begin to explore this intersection of the neuroscience and human-factors fields. To this end, the identification of a primary region of the brain for semantic comprehension, and an associated biomarker, serves as a starting point for this multi-discipline field.

To summarize these gaps, the following points can be highlighted:

- Current SA literature focuses primarily on stationary vehicle-based systems, and not man-portable wearable systems. It is unclear if the findings from these stationary systems continue to apply to wearable systems in a natural environment.

- Research into the brain dynamics of comprehension is complex, and the process has been identified throughout the brain depending on task. This shows a requirement for identifying a region of the brain and biomarker related to the specific task of semantic comprehension, which has yet to be identified.

Author	Year	Attribute	Stimulus Type	Classification Method	Accuracy
Simanova et al.	2010	’animal’ or ’tool’	Image	Bayesian logistic regression	83%
Maguire et al.	2010	Semantic or taxonomic relationship	Two spoken words	-	-
Chan et al.	2011	’living’ or ’non-living’	Written word	Support Vector Machine	64%
Murphy et al.	2011	’animal’ or ’tool’	Image	Support Vector Machine	72%
De Lucia et al.	2012	’natural’ or ’man-made’	Audio effect	Gaussian Mixture Model	56%
Costanzo et al.	2013	’living’ or ’non-living’	Image	-	-
Correia et al.	2015	10 words, presented in both Dutch and English, aiming to find language-independent concept	Spoken word	Linear SVM	54%
Behroozi et al.	2016	12 Categories, comprising living and non-living items (eg animal, flower, building)	Image	Naive Bayes	10%
Rekrut et al.	2020	5 categories (locational, actions, living, nonliving, numbers)	Written word	SVM / Random Forest / MLP	39.6% / 41% / 40.6%
Rekrut et al.	2021	5 categories (locational, actions, living, nonliving, numbers)	Imagined word	SVM / Random Forrest / MLP	39.1% / 41.9% / 40.7%
Sharma et al.	2024	’target’ or ’non-target’ item within a natural scene	Visual Image	SVM	83.6%
Yu et al.	2025	Semantically ’congruent’ or ’incongruent’	Written Word	CNN (EEGNet)	66.7% / 70% / 87.1%

Table 2.3: Attribute classification methods and accuracy from previous works

CHAPTER 3

LEVERAGING PREVIOUS USER EXPERIENCE FOR IMPROVED USER PERFORMANCE

3.1 Overview

The use of gaming, both as a method of training as well as inspiration for control methods has been found to have a positive impact on an individual's capabilities in adapting to a similar situation [151]. This, combined with the findings that sufficient training in a similar environment can lead to similar performance to those trained on a specific system [152], has led to the adoption of the use of video games as a method of simulation training [153]. From this, the question can be asked as to whether even if display designs tailored to a specific purpose should, in theory, perform better than displays commonly implemented in video games, does the experience using these less-than-ideal systems enable them to perform better?

3.2 Study Design

As usability testing is an effective method of evaluating a user interface, the choice was made to perform both quantitative evaluations to gauge the effectiveness of each design as an aide, as well as qualitative usability testing with experience gained through the quantitative task so that any potential issues or common thoughts regarding the designs could be highlighted. Additional care was made to ensure a large sample size was collected to avoid the issues associated with collecting a small sample size when performing usability testing. Additionally, using a quantitative method to provide user experience for the usability testing allowed us to collect quantitative data to assess performance and to collect the qualitative thoughts of participants to assess preference in a realistic use of the designs for their intended purpose.

3.2.1 Participants

This study was conducted through Qualtrics, a third-party online survey platform, with compensation and quality check service (to avoid repeat responses, partial responses, wrong file format uploaded etc.). Participants were screened to be within the target age group of 18-35 and were required to use a mouse to complete the survey.

A total of 103 responses were received, with 18 participants excluded from the analysis due to significantly poor performance with participants failing to notice stimuli above 75% of the total trials (n = 5), and reaction times with the time taken to point to the target and/or identify the symbols over three times the standard deviation to the mean (n = 13), giving a final sample size of N = 85.

To further investigate how game experience affects user performance in target localization and identification with four HUD designs, we divided participants into two groups by using the hours they spent on video game per week. People who spent 0-10 hours per week were casual gamers (n = 46), while those with 11-50 hours per week were experienced gamers (n = 36), with group definitions based on the State of Online Gaming Report 2020.

The study was approved by the local institute's research ethics committee, and participants were compensated for their time; any identifiable information was not included.

3.2.2 HUD Designs

Design decisions for the HUDs were made with the consideration that once a stimulus was in the field of vision, clarity of vision becomes an important factor; thus efforts were made when implementing the HUDs to ensure a less cluttered display when facing a stimulus. This design decision was also supported by game UI heuristics, which highlights the importance of providing "unobstructed views that are appropriate for the user's current actions" [62]. From this, the following HUD designs (Shown in Figure 3.1 were selected for comparison:

Radar - The use of radar as a localization method is extremely common, with a significant number of games making use of a permanent 2D overlay to indicate the direction of relevant aspects of the environment [154]. The radar consisted of a circular display with a fixed cone, which represented the participant's field of vision. When a stimulus was presented, an icon would appear on the radar along with the distance the stimulus was from the participant. As the participant rotated their 'head', the radar rotated similarly.

Indicators - As peripheral cues automatically aid in orientation towards the cue's location [48], emphasis was placed on including a design that used peripheral cues. As directional indicators are seen in a significant number of first-person video games [154], an indicator-based design was also selected. These indicators serve as an additional aide to the orientation towards the stimulus when the stimulus is out of the participant's field of view.

Compass - The use of a compass bar to aid with directional localization has seen implementations in reality, with AR implementations being designed as early as 1998 [69]. An additional secondary bar was provided at the bottom of the screen to allow 360-degree coverage of the environment, with the top bar representing information inside a user's field of view and the bottom bar representing information outside of the field of view. The icons were implemented so that if an icon went off one side of one of the bars, it would 'wrap around' to the other side of the opposite bar.

Ellipse - For the design representing a real-life implementation (Ellipse), the choice was made to draw inspiration from systems designed to be used in high-stress environments [55]. This is due to the security domain sharing a large number of traits with competitive video gaming, namely the ability to operate under high levels of stress as well as to quickly process and react to new information. It should be noted that whilst Ellipse appears similar some 'Radar' style HUD implementations, Ellipse does not make use of the centre of the HUD to display icons, with the icon being locked to the line of the ellipse, and the distance of the stimulus denoted by accompanying text. This differentiates it from radar where the icon distance from the centre varies based on the distance from the participant to the stimulus.

3.2.3 Protocol

After completing the screening, participants completed an online 2D interactive task to evaluate user performance and preferences with various first-person HUD designs (median completion time 31.5 minutes). After completing the task, participants were

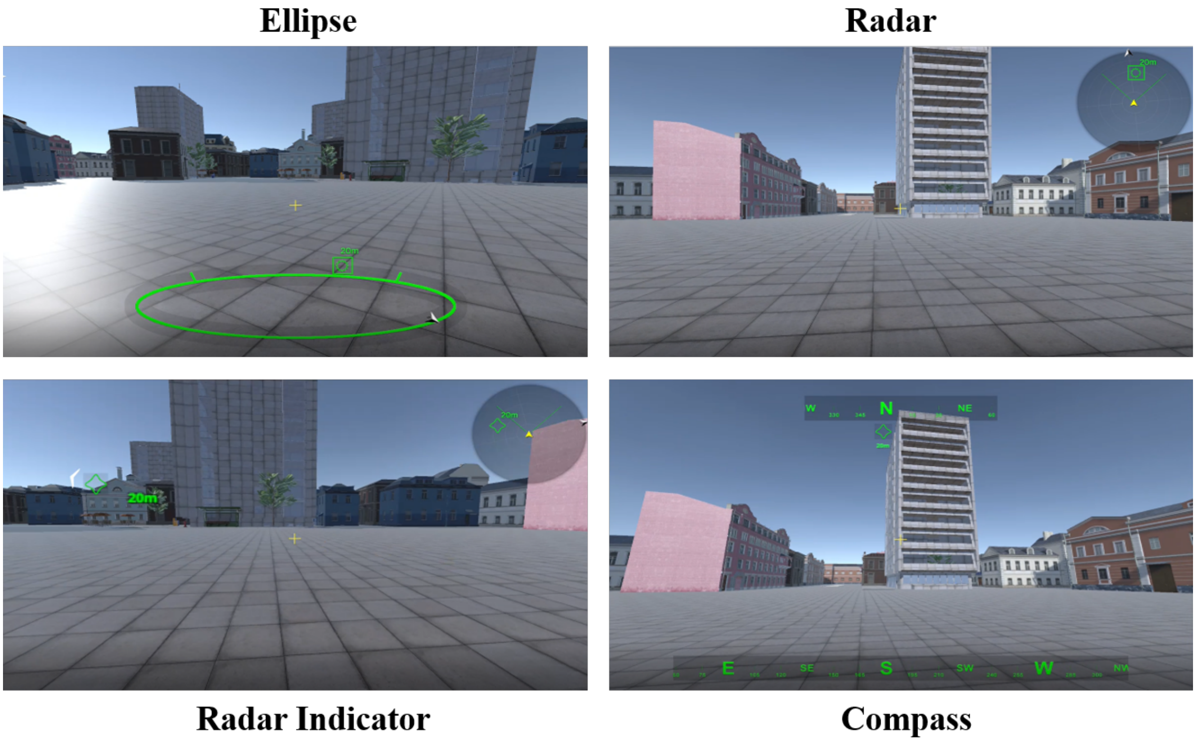


Figure 3.1: The four HUD designs used in this study

asked to respond to a set of qualitative questions for comparing user's preferences of the different HUD designs with direct pairwise comparisons.

3.2.3.1 Experiment Task

As the HUDs are designed to aid localization, the task paradigm was based on traditional sound localization paradigms [155, 156] to prevent any interference from using a visible stimulus to identify the location and to force the participant to rely solely on the HUD. This allowed the quantitative data to be used to evaluate performance as the HUDs were the sole guiding factor due to the lack of audio cues. To maintain similarity with traditional paradigms, the participant indicated the direction of the stimulus by turning their 'head' towards the stimulus, which served as the reference point and direction for all the HUDs. A reticule in the centre of the screen mitigated the potential inaccuracy from using head-based direction indication [156].

Additionally, a realistic virtual environment was used as it has been shown to aid localization accuracy and reduce confusion[155]. To account for individual participant hardware, the ability to modify mouse-sensitivity was provided as well as steps being

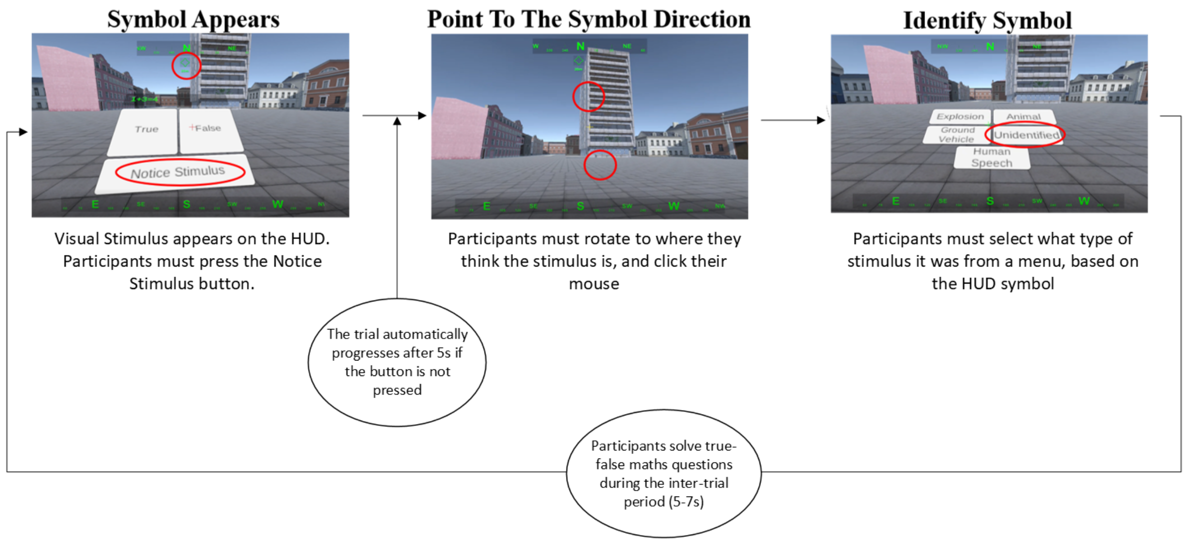


Figure 3.2: Diagram detailing the protocol of a single trial

taken to ensure the image on the screen appeared as the same physical size across participants. Instruction slides were provided to the participant preceding every step of the experiment. The symbols were designed so that potential resolution and fidelity issues from transference to XR were accounted for.

Participants were trained on the symbols before beginning any task training. To further aid the ability to distinguish features of the symbols, as well as account for potential resolution and fidelity issues of AR, efforts were made to ensure the symbols were simplistic as well as quickly and easily identifiable. To allow a sufficient level of understanding regarding the symbols, before beginning any task training, the participant was trained on the five symbols that could appear on the displays.

Each condition consisted of 40 stimuli, representing the five symbols in each of the four cardinal directions, repeated once. When a stimulus was presented, its direction was modified by ± 10 degrees to ensure the participant could not rely on muscle memory when locating the stimulus. To account for learning, the order of the HUD conditions and of stimuli were randomized. To aid the participant in identifying the direction they were indicating, as well as the stage they were currently in, a colour-coded reticule was provided, consistent across designs. The participant was also required to attend to a simple true-false maths task. This task served as a method of ensuring the user's attention was not constantly placed on the HUD, requiring them to either consciously attend to the HUD or notice a stimulus through their peripheral vision. A diagram detailing this paradigm can be seen in Figure 3.2.

Calibration & Training - To ensure that participants understood the trial procedure without the possibility of observer intervention, calibration and training stages were implemented. Calibration consisted of no HUD, with 8 easily visible stimuli being presented in a clockwise direction. The participant then aligned the centre of the camera, indicated with a reticule, with the sphere and clicked to 'point' at the sphere. Training consisted of trials similar to the main task, though with access to a feedback screen showing metrics related to their performance, as well as instructional prompts in the event they failed to perform a stage correctly. One of the training-only prompts would not allow a participant to continue past the pointing stage if their indicated direction was more than 5 degrees off the true stimulus position - This served as a method of training the participant where the symbol would be on the HUD when they were pointing directly towards it.

Localization & Identification Task - The localization and identification task was comprised of three stages, representing noticing, locating, and identifying a stimulus. After 5-7 seconds of the previous trial's completion, a stimulus was presented. The participant then had to indicate their awareness of a stimulus by pressing a button located within the environment when they noticed the stimulus. In order to ensure the completion of the task, an automatic timer of 5 seconds on noticing each stimulus was implemented, with failure to notice a stimulus within that time having the task auto-proceed to the pointing stage. When the stage progressed past noticing, either through the time-out or a participant noticing the stimulus, the maths task was disabled to act as an additional indicator that the participant was to locate the stimulus. Secondly, the participant turned to directly face the stimulus, pressing the left mouse button when they believed they had positioned correctly. Finally, the participant turned back to north and identified the meaning behind the symbol by clicking the button with the word that the symbol represented from a list of buttons with every possible meaning. The experiment then re-enabled the maths task for them to attend to until the next stimulus was presented. During training, the meaning buttons had the relevant symbol next to them to provide additional training to the participant, however, these guides were disabled during the test condition.

3.2.3.2 Questionnaires

VICER: The VICER questionnaire is a four-point rating scale system devised by industrial collaborators for this study, and aims to assess individual user preferences as well as identify possible advantages or disadvantages of each HUD design from five perspectives of visualization (visibility, conspicuity, emphasis) and interpretation (readability, interpretability) [157, 158]. The five keys aspects used to assess the 2D HUD designs were as follows, with all participants ranking the designs with "1" being the most preferred option and "4" as the least preferred option.

- **Visibility (V)** - The ability to see and distinguish the display elements overlaid in the environment.
- **Interpretability (I)**- The ability to understand the content of what the display was communicating or how to use the information the
- **Conspicuity (C)** - Whether the display attract the appropriate amount of the attention.
- **Emphasis (E)** - Whether the display place the right emphasis on the symbol/target information.
- **Readability (R)** - The ability to recognize symbol/target and pairing information quickly and accurately.

Pairwise Comparison: Pairwise comparison provides a direct comparison on which design the user preferred along with qualitative feedback to provide possible reasoning as to why a design was preferred over another. Participants were asked to choose their preferred HUD design when comparing two designs and then provide the reasons for their selection. Examples given included, but were not limited to, the following: is A or B clearly indicating the target, or does A or B block the background more, etc.

The survey closed with the opportunity to provide additional comments about the 2D task or the survey. All participants completed the survey questions in the same order.

3.3 Data Analysis

Of the 103 total responses, 18 participants were excluded due to significantly poor performance (Failing to notice more than 75% of trials, N=5), and other behavioural metrics being greater than 3 S.D. from the mean (N=13), giving a final sample size of N = 85. To further investigate how previous experience affects performance, participants who spent 0-10 hours per week on video games were classified as Casual Gamers (N = 46), with those having 11-50 hours per week classed as Experienced Gamers (N = 36).

Any participants with more than 50 hours per week were not included in this analysis (N = 3).

To analyse behavioural performance, three main metrics were assessed: Time to locate the stimulus (point time), time to identify the category of stimulus (identify time), and the angle error in localization between the participants' indicated direction and the true direction. To further analyse the change in performance over time, linear regression was conducted on the time taken to identify stimuli. For qualitative results, participant rankings based on chosen qualities (VICER) was investigated to gauge HUD usability and preference.

All statistical analysis was conducted in SPSS Statistics 27.0. Distribution normality was assessed with a Shapiro-Wilk Test, then parametric analysis was conducted with one-way ANOVA and Tukey post-hoc. Non-parametric analysis was performed with Friedman and Wilcoxon signed-rank test. In all tests, significance level α was set to 0.05, and effect sizes are reported using partial eta square (partial η^2).

3.4 Results

3.4.1 Behavioural Results

3.4.1.1 Performance Metrics

For the behavioural metrics, each of the four displays was measured on Point Time, Identification Time, and Angle Error. A Shapiro-Wilk Test showed significant departure from normal distribution of all three metrics, with Point Time (W(340) = 0.881, p = 0.000), Identification time (W(340) = 0.913, p = 0.000), and Angle Error (W(340) = 0.71, p = 0.000), respectively.

For pointing times, the mean results are as follows: Ellipse = 1.913; Radar = 1.732; Radar-Indicator = 1.938; Compass = 2.259. A Wilcoxon Signed-Rank test reveals that the Compass was statistically significantly different to the other three conditions in regards to Point Time, with statistics as follows: Ellipse (Z = -2.592, p = 0.010); Radar (Z = -3.784, p = 0.000); Radar Indicator (Z = -2.903, p = 0.004). Overall, HUD type was found to have a significant main effect on Point Time ($\chi^2(3) = 16.355$, p = 0.001.).

Identification Time revealed no significant main effect of HUD type ($\chi^2(3) = 6.812$, p = 0.078).

For absolute Angle Error, the mean results, in degrees, are as follows: Ellipse = 3.989°; Radar = 6.753°; Radar-Indicator = 8.370°; Compass = 88.305°. For statistical significance,

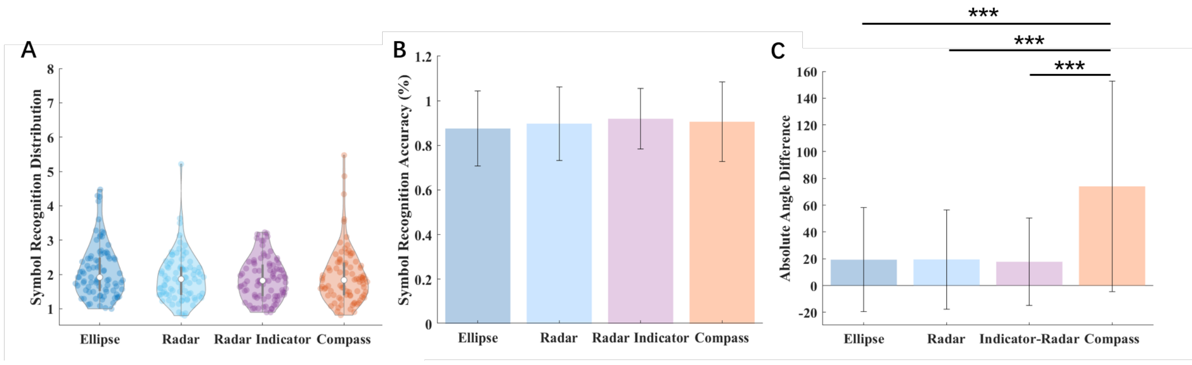


Figure 3.3: Behavioural Results for HUD Design Study (Reproduced from [159])

a Wilcoxon Signed-Rank test showed that Compass was significantly different compared to the other three conditions, with statistics as follows: Ellipse (Z = -5.896, $p < 0.001$); Radar(Z = -5.760, $p < 0.001$); Radar Indicator (Z = -5.887, $p < 0.001$). Overall, HUD type showed to have a significant effect on absolute Angle Error in regards to localizing the stimuli ($\chi^2(3) = 46.962$, $p < 0.001$).

A visualisation of these results can be seen in Figure 3.3.

3.4.1.2 Effect of Experience on Performance

A main effect of game experience was found in the following conditions: Ellipse (F (1,80)= 7.486, p = 0.008, partial η^2 = 0.086); Radar (F (1,80) = 5.765, p =0.019, partial η^2 = 0.067); and Radar Indicator (F (1,80) = 4.605, p =0.035, partial η^2 = 0.054), with the effect that target Identification time was higher in the Casual group than the Experienced group, while there was no group difference in Compass (F (1,80) = 0.54, p = 0.465, partial η^2 =0.007). To further explore the tendency changes with time, a simple linear regression was calculated. A significant regression equation was found in the Experienced group for Ellipse (F (1,8) = 9.309, p = 0.016), with an R^2 = 0:538; Radar Indicator (F (1,8) = 9.241, p = 0.016), with an R^2 = 0:536; and Compass (F (1,8) = 9.860, p = 0.014), with an R^2 = 0:552 compared to the Casual group. These effects were not shown in the Radar condition. A visualisation of the regression can be seen in Figure 3.4.

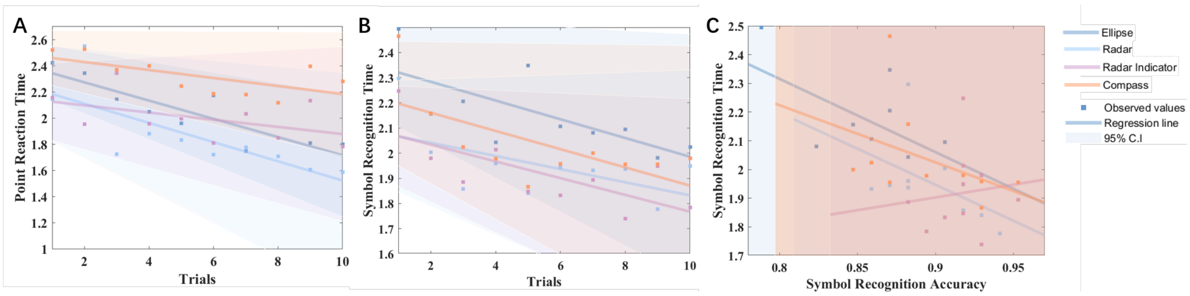


Figure 3.4: Linear Regression of Performance Improvement for HUD Design Study (Reproduced from [159])

3.4.2 Questionnaire Results

3.4.2.1 VICER Results

To investigate participants' preferences of each display, The total number of participants' ranking distribution for the four HUD designs were compared, with "1" as the most preferred option and "4" as the least preferred option. The results show that more than 30 subjects choose "4" (Least Preferred) for Compass (43%) from Visibility (n = 38), Interpretability (n = 34), Conspicuity (n = 31), Emphasis (n = 34), and Readability (n = 39), and more than 24 subjects choose "1" (Most Preferred) for Radar Indicator (35%) from Visibility (n = 24), Interpretability (n = 29), Conspicuity (n = 28), Emphasis (n = 35), and Readability (n = 26). These results have been visualised in Figure 3.5.

3.4.2.2 Pairwise Results

To further identify participants' preferences, the pairwise comparison was analysed by calculating the preferential percentage for each HUD display. Pairwise comparison results showed that 33% of the participants chose Radar Indicator as their preferred display, with reasons given highlighting the indicator helping with locating and identifying targets. 27% of the participants chose Radar as their preferred display. Whilst 21% of the participants chose Ellipse, with reasoning given that it was easy to understand. Another 21% of participants chose Compass, while people who disliked Compass gave reasons citing showing too much information and causing confusion.

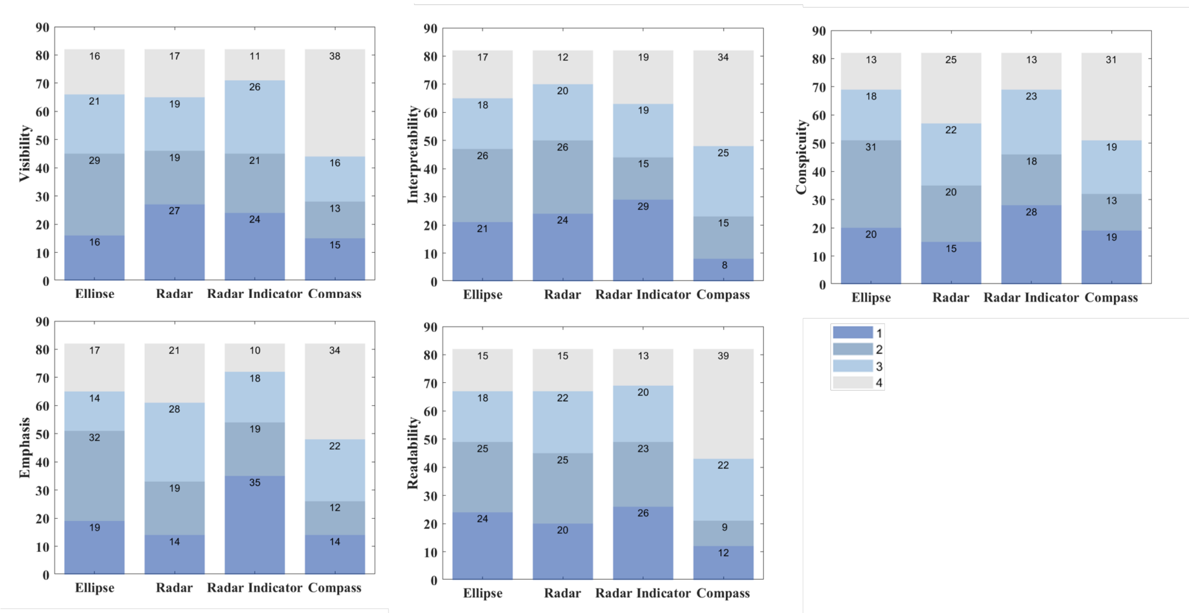


Figure 3.5: VICER Results for HUD Design Study (Reproduced from [159])

3.5 Discussion

By conducting an online survey with online interactive 2D game, it was possible to observe a set of general comparisons for the HUD displays, with Radar Indicator showing an overall higher subjective preference, while Radar, Ellipse, and Compass followed as less preferred, even when most HUD designs shared a similar behavioural performance level.

3.5.1 Comparable Localization Performance between Common Video Game and Specialized HUD Designs

The results indicated no statistically significant difference between the different HUD displays in target localization and identification performance. This suggests that video game-inspired HUD designs are comparable in performance to specialized HUD designs for this purpose, which may indicate that commonly seen video game HUD designs have the potential to perform as well as industry-implemented HUD designs for specific usage in fields where physiological factors become relevant. These analogous results may be due to similar designs of Ellipse, Radar, and Radar Indicator, as these three displays were all egocentrically designed with a similar circle ring rotating as users turned to locate the targets. This similarity of Ellipse with Radar made it easier to understand and use even though it was not implemented for common use, as the circle ring was straightforward, showing directions and symbols, even with the possibility that participants had not encountered it before this study. The main difference of these three designs was the position of the circle ring, with Ellipse on the bottom, while Radar was on the top right corner, which did not lead to performance difference but revealed subjective preferences.

However, as the outlier design, Compass showed the worst performance in localizing targets accuracy with a significantly higher angle difference, though the identification time was within the normal range as with the other three displays. This could be explained that, firstly, compass might be not appropriate for directional discrimination, especially front and back, given it was implemented with two bars (top and bottom) for more direction information delivered, which might have caused the complication that participants were confused by using the wrong bar; secondly, even though Compass was designed as a common video game HUD, it is still less common overall within video game design, and therefore it is possible that participants had less overall experience with a dual-compass design and displayed decreased performance due to this.

3.5.2 The Impact of Video Game Experience on HUD Usability

Comparing participants' performance with different amounts of game experience, improved target localization and identification performance was found in participants that spent more hours a week (> 10h) engaging in video games, and could be seen as experienced gamers, while participants with less game experience presented an expected lower reaction time on target identification.

As shown in previous research, a possible reasoning for the improved performance for experienced gamers is the finding that the cognitive profile of individuals who regularly engaged in video games showed improved cognitive performance in visual and attention capability [160], enhanced cognitive processing speed [161], and improved working memory [162, 163].

In this study, as an online 2D interactive task, individuals were asked to participate in it by using their mouse to target the symbols as a virtual pointer. Experienced gamers may have found greater familiarity with using a mouse as the primary user interface. Hence, the findings may not fully correlate with previous studies that compared the virtual and physical pointing performance, as the act of physically pointing may produce different results [164, 165].

In addition, a learning/training effect was found in both traditional and specialized HUDs as use-time increased. More specifically, only the specialized HUD (Ellipse) revealed a negative correlation between symbol accuracy and symbol identification time over trials. This negative tendency may indicate that as Ellipse was an unfamiliar design for both casual and experienced gamers, it might take more time and practice to reach a similar performance as common traditional displays. This may indicate that game experience could be seen as a critical factor with regards to the comparable performance. This leads to a key recommendation that future HUD designs be more mechanically and/or heuristically similar to video game HUD designs to allow for the transference of familiarity and previous experiences to increase user performance.

3.5.3 Subjective Preferences Difference between Common Video Game HUD and Specific HUD Designs

Participants tended to select Radar Indicator as their preferred display in terms of subjective satisfaction evaluations and direct pairwise comparisons, although there was no significant difference in behavioural performance between common video game and specialized HUDs. This might indicate that even though individuals may tend to prefer some specific designs subjectively, their objective performance may not differ that much. The reasons for these inconsistencies and factors that might affect participants preferences could be listed as follows:

Radar Indicator: Qualitative evaluation of both VICER and direct pairwise comparison revealed Radar Indicator as the most satisfactory display with the highest preference, though quantitative performance did not show any difference. As described by participants, specific reasons given include examples such as: *"Radar Indicator is better because it visually shows where the object is in the environment", "Radar Indicator utilizes the mini-map to track the target, but it also shows an on-screen indicator"*. This opens the possibility that participants were more familiar with Radar due to its more commonplace usage, when compared to the other three displays. Additionally, Indicator provided additional assistance in understanding how to use it to locate and identify the targets. Since the use of Radar-based HUDs is quite common, especially in First Person games, it is not surprising that adding Indicators to Radar makes it easier and direct to localize the target symbol [166].

Ellipse: Ellipse showed marginally equal evaluation as Radar in both VICER and direct pairwise comparison, which may indicate that participants gave a similar response to Radar and Ellipse. First, as Ellipse and Radar shared a common circle ring, though with different positions, it may lead to locating targets using a similar method. The reasons given for participants' selection between Ellipse and Radar for the target localization and identification aides were shown by comments such as: *"The Ellipse circle makes it easy to determine what you are doing, specifically the green colour makes it very easy", "I preferred Radar over Ellipse because the location of the symbol is marked in the 3D space in front of me, which is the easiest to locate, instead of a 360 degree space at the bottom of the picture"*. Then, as Ellipse was derived from a real-life implementation which is used in stressful and night vision circumstances, it could be difficult for individuals to get used to it at the beginning [167], however performance could be improved through practice and learning, as seen in Figure 3.4.

Compass: Compass showed the lowest satisfaction in both VICER and pairwise comparison, which also corresponded to target localization performance with the highest localized error. Although Compass was drawn from traditional first-person role-playing games, it was altered to have two compass bars located both on the top and bottom of the screen to show the orientation information, which may have caused confusion of the exact facing in regards to the front and back. Participants' feedback had comments such as: *"The setup of Compass confused me a lot as I didn‚Äôt know which direction to look", "I prefer Compass, because it's easier to use once you understand the co-ordinates", "Compass was confusing at times with the two different compasses, I often messed up which one I needed to focus on and look at"*. This decrease in performance and confusion regarding direction could potentially be caused by the fact that the participants were recruited from the general public, and may not have the prior experience needed to make successful use of the Compass, as opposed to professional or expert users who may be more familiar with the design [168].

3.6 Summary

As seen in this study, common HUD designs found in video games can not only rival bespoke designs in performance, but potentially provide a greater benefit to the user. Additionally, we found that participants preferred designs they had more experience with, as seen in Eclipse/Radar where there was an equal performance, but a greater preference for the common design.

3.7 Limitations & Future Work

Whilst the data regarding these four display designs is comprehensive, there are more than four methods of displaying this data. This means that whilst it can be determined which of these designs are superior in certain aspects, there is no guarantee that there is not a better design available among more specialized or unique display methods. Additionally, the method in which the scenario was deployed (That of a 2D web-based task) has potential implications when transferred from the limited field of view of a screen to the larger field of view as afforded by Virtual/Augmented Reality headsets, as with a larger display space available, some of the designs that were often described as conspicuous or as taking up the entire screen may prove to be more suited to a larger field of view. In regards to the specific displays, training for each display was a minimum of 5 trials, so it is possible that with sufficient training one design may prove superior than another that seemed better with minimal training, as the results provided here are for minimal training on the displays, with the participant performing only 45 trials per display type. Finally, task acclimatization is a possibility, though amongst participants efforts were made to ensure that each design appeared equally in each possible position of order to balance the effects of acclimatization through all the different designs.

As seen in this study, commonly implemented HUD designs have the capability to rival bespoke designs, as well as previous studies demonstrating common control methods can also rival specialized designs. This raises the question as to whether this trend is true across a broad spectrum of environments and applications. Further research could also be done into the capabilities of these displays under a variety of factors that an individual requiring use of an SA-aide might encounter, such as fatigue, extreme time pressure, or stress, allowing further insight into whether common implementations are affected more by negative SA-factors than the bespoke designs. Finally, additional research could be done into the display designs themselves, as a more bespoke display system may prove superior to common designs provided enough training is provided on the bespoke design, or perhaps a specific demographic of user performs better with a specific style of display.

CHAPTER 4

How Physiological Stress affects HUD use and SA for Target Localization and Identification

4.1 Overview

The ability to detect and locate audio sources that cannot be visually seen forms a core concept in almost every aspect of life, especially in stressful or dangerous situations where mistakes can lead to injury or even fatality. Therefore, we see large efforts in improving localization capabilities through various means such as Heads-Up-Displays (HUDs). However, these HUDs are often complex and require training to be proficient with them, especially when under stressors from a dangerous environment. In order to avoid risks in training, virtual reality is often used to simulate risk-filled environments, however, little research has been done on the effectiveness of HUDs in a state of stress from these environments, and whether HUDs continue to provide tangible benefits in a state of stress from simulated risk. To answer this gap in training knowledge, this chapter presents the findings of a study where 20 participants were placed into a 'dangerous' virtual environment whilst being physically elevated to induce a state of heightened stress. Behavioural results of a complex target localization and identification task show the presented HUD continues providing significant benefits to behavioural performance in target localization and identification even when subjected to real-world stress factors, as well as identifying stress as a significant factor in performance level.

4.2 Methodology

In order to safely place participants in an environment that both mimicked reality, as well as to increase the stress participants further felt, the localization task was performed in a virtual environment, developed in Unity (v2019.3.3f1), with the participants being placed into two urban environments, one on ground level, and another on a plank leading off the edge of a skyscraper.

4.2.1 Apparatus

Participants wore a HTC Vive Eye Pro (2 x 1440 x 1600 resolution, 90Hz refresh rate, 3D spatial sound, and 120 Hz eye tracker sampling rate) for the duration of the experiment. To further ensure the auditory stimuli were properly spatialized on the HTC Vive's built-in stereo headphones, the official Vive 3D Sound Spatializer Plugin (version 1.2.0) was used, as well, as a pilot study was conducted to verify the spatializer was functioning correctly. The spatializer made use of the default Head Related Transfer Function provided with the spatialzer plugin. To further increase participant immersion, participants wore HTC Vive Trackers on their feet, with virtual feet being projected into the virtual environment to aid with traversing the plank safely and effectively by synchronizing feet location between the virtual and real environments.

4.2.2 Participants

The study consisted of 20 healthy right-handed participants (Mean age 26.84

¬±

3.01; Range 21-35 years old, 6 Female), all with normal or corrected-to-normal visual and auditory function. No participants reported any history of psychiatric disorders, neurological diseases, or alcohol or drug abuse. Each participant received 60 AUD for their participation. Before the experiment, informed written consent was obtained from all participants. The Institutional Human Research Ethics Committee approved this study protocol involving human subjects (ID Redacted for Review)

4.2.3 Experiment Procedure

Participants were required to complete an audio/audiovisual target localization and identification task in two stress conditions, with the two conditions featuring the par-

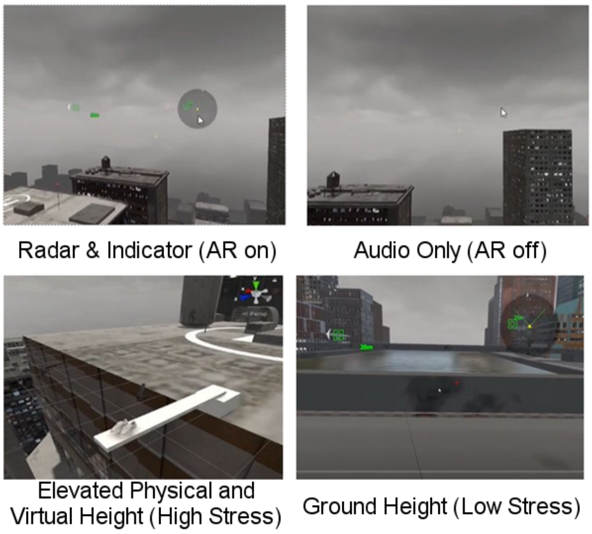


Figure 4.1: Representation of the four conditions in the Physiological Stress study

ticipants performing the experiment at different physical heights to invoke different stress levels, as shown by previous works. In these two conditions, some trials were aided by a heads-up display, and some were unaided, leading to four total conditions (4.1). This results in a 2-by-2 within-subject design, with two independent variables of HUD guidance and stress level. Stress conditions were counterbalanced across participants, with half completing the high-stress condition first, and the other half completing the low-stress condition first. HUD Guidance was counterbalanced by having every trial combination (Direction, category) having a HUD-assisted and unassisted trial. Trials were presented in a randomized order as to prevent learning related to the order of symbols, pre-guessing locations, or assuming HUD support. In total, participants completed 80 trials per condition, separated into 4 blocks of 20 trials. The complete experiment took approximately two hours, with approximately 45 minutes for each condition, 20 minutes of training, and a 10-minute break between stress conditions with a calming breathing exercise as rest. The training consisted of a calibration stage, where participants had to point at 8 visible spheres that appeared in a clockwise direction to ensure that participants comprehended the localization aspect of the study, followed by 40 training trials that formed the training block.

Participants received training on the experiment paradigm, being required to complete both a HUD-symbol pairing task to ensure understanding of the symbols, as well as to complete a block of 20 trials as a training block. Between each block, participants were required to answer questionnaires regarding their current emotional state, as well as be-

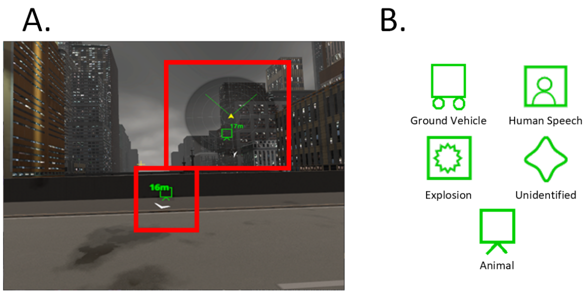


Figure 4.2: A) HUD elements used to indicated stimulus. B) All possible symbols that could appear on the HUD.

ing provided a rest break to counteract fatigue during the experiment. Participants were also required to complete a relaxation task consisting of breathing exercises between conditions to ensure that each participant was calm when beginning each condition.

4.2.3.1 Stimuli

The study contained a background noise of city sounds at 60db, screened as not to contain any sounds of stimuli categories, with stimuli being 80% of the background volume so as not to be immediately distinguishable. The length of each auditory stimulus was approximately one second in length for the audible portion of the sound files. Stimuli were one of five categories: Vehicle, Human Speech, Animal, Explosion, and Unknown. These categories were chosen as they are sounds that can be potentially obstructed by the environment, fulfilling the condition of being a non-line-of-sight stimulus. An example of the HUD, as well as the category symbols, can be seen in figure 4.2. Trials were generated so each category was presented 4 times (2 Guided, 2 Non-Guided) in each cardinal direction within an arc of 30 degrees each side to prevent muscle memory for the directions.

4.2.3.2 Localization & Identification Task

The localization and identification task was composed of three stages, representing the processes of Noticing, Locating, and then Identifying a stimulus. Participants completed all of these stages using Vive Pro controllers that were represented in the virtual environment (4.3).

For the 'Notice' stage, an auditory stimulus was presented at a random delay between 5-7 seconds after the previous trial ended. In HUD-guided trials, the HUD denoted both

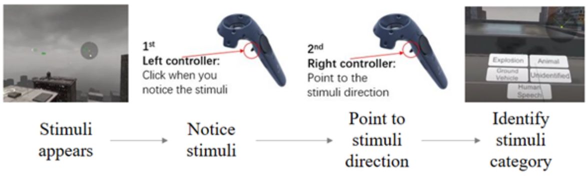


Figure 4.3: The protocol for a single trial in the Stress Experiment

the category of stimulus as well as the direction. Participants were then required to press a button on the VR controller to indicate their awareness of a stimulus. If the participant failed to notice a stimulus, the trial automatically progressed to the next stage with a message informing the participant that they had missed a stimulus and had to attempt to locate the stimulus.

In the 'Locate' stage, participants were required to point to the location they believed the stimulus to be. To ensure participants knew where they were pointing, the controller within the VE featured a laser pointer to aid participants in the localization task. The verticality of where the participants pointed was not considered, and during analysis, height was normalized to be on a 2D plane around the participant.

In the 'Identify' stage, participants were required to return to the starting orientation and select an option from a menu in the VE relating to what category of the stimulus was presented. As the HUD marker for the stimulus was removed on completion of the Locate stage, participants were not able to refer to the HUD during this stage, instead having to remember either the symbol or the sound without any visual aids.

4.2.4 Inducing Real-World Stress in VR

In order to induce a state of stress in participants whilst within the VE, participants were asked to walk to the middle point of a 2-metre-long wooden plank before beginning each block of trials. As demonstrated to evoke significant differences in stress in previous works, for one condition, the plank was placed on the ground, and the VE had the participant standing on the ground, whilst for the other condition the plank was placed 0.65m above the ground, with the virtual environment having the plank leading off the edge of a skyscraper (Virtual Height = 250m). This subjected participants to differing levels of stress between the two conditions. To ensure participant safety, participants were required to wear a safety harness attached to an Australian Height Safety Services Fall Prevention System. To prevent the stress from the safety harness being a confounding factor, participants were asked to wear the harness for both conditions.

To measure participant stress immediately after each block, the participants were asked to indicate their subjective opinion on their own state using the Self Assessment Manikin, which asks participants to rate their current Valence, Arousal, and Dominance, and allows an insight into the participant's subjective stress level by rating their level as one of five reference states, or between two states, totalling 9 possible ratings[169]. In the self-assessment Manikin provided to participants, For Valence, 1 represented Happy and 9 represented Unhappy. For Arousal, 1 represented Excited and 9 represented Calm. For Dominance, 1 represented Controlled, and 9 represented In Control[170]. Measuring the SAM after each block of trials, including the training block, gave a total of 5 measurements of emotional state per condition.

4.2.5 Data Processing & Analysis

For behavioural data, the following were assessed: Participant reaction times during each stage, the success rate for stages where failure was a possibility (whether each stage was successfully completed), and absolute angle difference in localization (The absolute angle difference between where participants pointed in a flat plane around them and the true direction of the target stimulus in the same flat plane). Participants' reports on the Self-Assessment Manikin between trial blocks were also assessed to ensure stress was a significant factor during the experiment.

The data was processed in MATLAB v2020b using the MATLAB Statistics Toolbox, and was separated into four conditions: Low-Stress HUD-On, High-Stress HUD-On, Low-Stress HUD-Off, and High-Stress HUD-Off. Any outliers of more than 3 times the SD were removed before analysis of each stage. A Shapiro-Wilk Test was applied to test the normality of distribution, then an N-Way ANOVA was conducted, with significance level α set to 0.05. In order to correct for α inflation, a Bonferroni correction was applied in post. For behavioural data, the statistical tests compared the four conditions in a pairwise manner to determine significance. Based on G*Power [171], to achieve a power of 0.90 with an alpha level of 0.05, 20 participants should result in an anticipated effect size of 0.38 with a critical F value of 4.41.

In all figures, * = $p<0.05$, ** = $p<0.01$, and *** = $p<0.001$. Means are reported in the format of Mean$\pm$SE. Error bars on all figures are representative of SE.

To account for cybersickness, participants had to complete an SSQ after each condition. Participants averaged 2.79 points higher on the SSQ related to the stress condition. Significant categories were 'general discomfort', 'difficulty focusing', 'salivation', and 'dif-

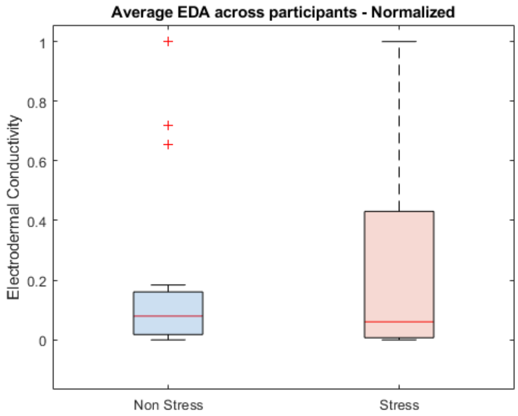


Figure 4.4: Normalized Mean Electrodermal Activity for each participant during each condition

ficulty concentrating', which could be due to the stress of the environment. The average for the stress condition was 5.47, and the average for the non-stress condition was 2.68.

4.3 Results

4.3.1 Verifying Stress Levels

4.3.1.1 Electrodermal Activity

In order to measure participant stress levels, participants' electrodermal activity (EDA) was recording using an Empatica E4 wristband throughout the experiment. To calculate the participant's mean EDA, the EDA was taken during the performance of each block of trials, discounting EDA from the inter-block rests, with this data then being normalized to remove variance between participants. The mean of these data points was then taken to calculate a mean EDA during the active completion of the experiment. The IQR for EDA for each condition is as follows: Non-Stress = 0.02-0.16, Stress = 0.01-0.43. Non-Stress additionally had three outliers, indicating a significantly higher EDA when compared to the rest of the participant cohort. This difference in the normalized EDA can be seen in Figure 4.4

4.3.1.2 NASA-TLX

After both conditions, participants were asked to indicate the perceived workload of the tasks on the 11 Point TLX Scale [172]. The mean scores for each category across participants can be found in table 4.1. Significance can be seen in four categories, with participants rating the stress condition significantly higher in the Mental, Physical, Effort, and Frustration categories. This increase indicates an increased workload placed upon participants when performing the task in the Stress condition.

NASA-TLX Category	Stress Mean Score	Non-Stress Mean Score	Significance
Mental	5.2 ± 2.25	3.95 ± 1.56	** ∣ P = 0.007
Physical	5.15 ± 1.59	3.25 ± 1.81	** ∣ P = 0.004
Temporal	3.2 ± 1.99	2.55 ± 1.32	NS ∣ P = 0.07
Performance	3.75 ± 1.34	3.5 ± 1.53	NS ∣ P = 0.24
Effort	5.7 ± 1,82	3.7 ± 1.85	*** ∣ P <0.001
Frustration	3.75 ± 2.12	1.9 ± 1.14	*** ∣ P <0.001

Table 4.1: NASA-TLX mean scores across participants for each condition ∣ ** = P < 0.01 ∣ *** = P < 0.001

4.3.1.3 Self Assessment Manikin

To ensure the effectiveness of the virtual environments on invoking stress within participants, the participants' ratings on the Self Assessment Manikin scale were investigated (4.5).

Valence: In regards to Valence (How happy or unhappy the participants felt), no significant differences were found between the two conditions, however, valence was consistently higher for the high-stress condition across participants (Participants were unhappier), with the means for High-Stress being 4.05, 4.3, 3.95, 3.70 during the breaks, and the means for Low-Stress being 3.15, 3.30, 3.55, 3.40 respectively. *Arousal:* In regards to Arousal (How excited the participants felt), participants were required to rate themselves, with lower values signifying higher excitement levels. A significant difference was found between the two conditions during the second break, with participants feeling significantly more excited in the high-stress condition (p = 0.035). Arousal was also constantly lower for the high-stress condition across participants (Participants were more excited), with the means for High-Stress being 5.05, 5.30, 5.70, 6.30 during the breaks, and the means for Low-Stress being 6.10, 6.65, 6.45, 6.60 respectively. *Dominance:* In regards to Dominance (How in control of the situation the participants felt), a significant difference was found in all but the final block of trials, with participants feeling significantly less in control of the High-Stress condition during the first three blocks of trials (p = 0.011, 0.025, 0.041 respectively). Dominance was consistently lower

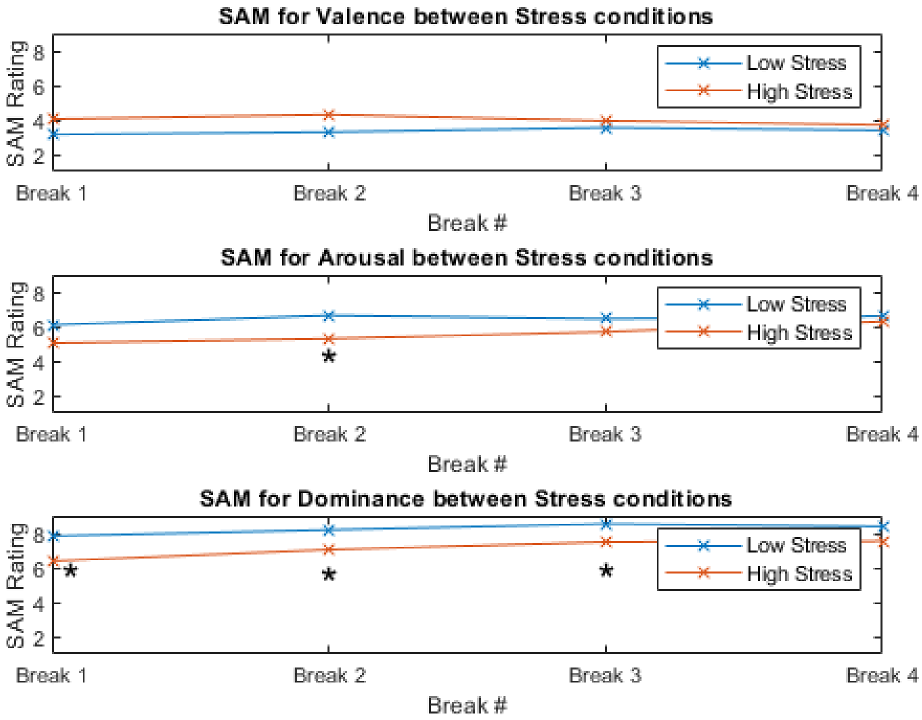


Figure 4.5: Average response on the Self Assessment Manikin during inter-block breaks

for the high-stress condition across participants (Participants felt less in control of the situation), with the means for High-Stress being 6.40, 7.05, 7.50, 7.55 during the breaks, and the means for Low-Stress being 7.85, 8.20, 8.55, 8.40 respectively.

4.3.2 Localization and Identification Performance

To investigate the performance of the HUD under the effects of stress, participant reaction times in all stages, as well as the accuracy of all stages, were explored. This allows the effects of both stress as well as the HUD to be observed in the target localization and identification task. Additional analysis was performed on survey data collected during the inter-block breaks to ensure that the stress condition had a significant effect on other expected metrics.

4.3.2.1 Notice Stage

To investigate the time taken for participants to indicate awareness of stimulus presentation (4.6), N-way ANOVA indicated a significant difference in time taken to indicate

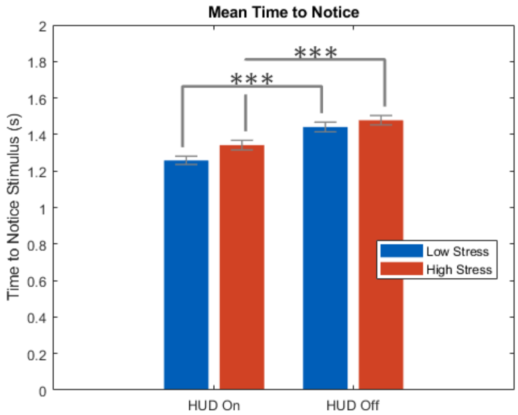


Figure 4.6: Time taken to indicate awareness of presented stimuli - Stress Experiment

between the HUD-on and HUD-off conditions ($p < 0.001$ for both Low- and High- Stress), with the effects of HUD guidance overall being found to be significant ($F(1,2998) = 39.77$, $p < 0.001$) No significant difference was found for the effects of stress in pairwise comparisons of HUD conditions, however the effect of stress overall was found to be significant ($F(1,2998) = 5.67$, $p = 0.017$). For the mean time to notice presentation of a stimulus, in the HUD-on condition the mean time for low-stress was 1.26 ± 0.02, and the mean time for high-stress was 1.34 ± 0.03. In the HUD-off condition, the mean time for low-stress was 1.44 ± 0.03, and the mean time for high-stress was 1.48 ± 0.03.

For participants' ability to successfully indicate awareness of stimulus presentation, N-Way ANOVA indicated no significant difference in success rate between the HUD-on and HUD-off conditions, with the effects of HUD guidance overall being non-significant ($F(1,79) = 3.27$, $p = 0.075$), furthermore, no significant difference was found for the effects of stress in both HUD conditions (4.7). For the mean percentage of successful stimulus awareness in the HUD-on condition, the rate of success for low-stress was $97.59 \pm 1.50\%$, whereas the rate of success for high-stress was $97.62 \pm 1.14\%$. In the HUD-off condition, the rate of success for low-stress was $94.20 \pm 2.44\%$, and the rate of success for high-stress was $94.13 \pm 2.25\%$.

4.3.2.2 Locate Stage

In regards to participants' ability to indicate the location of the stimuli, the time taken to locate the stimuli was investigated (4.8). N-way ANOVA of all trials indicated significant differences in the time taken to locate stimuli between the HUD-on and HUD-off conditions ($p = 0.0011$ and $p = 0.025$ for Low- and High- Stress respectively), with HUD

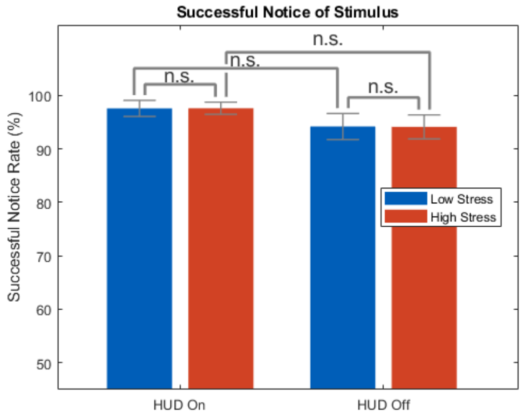


Figure 4.7: Success Rate for noticing the presented stimuli - Stress Experiment

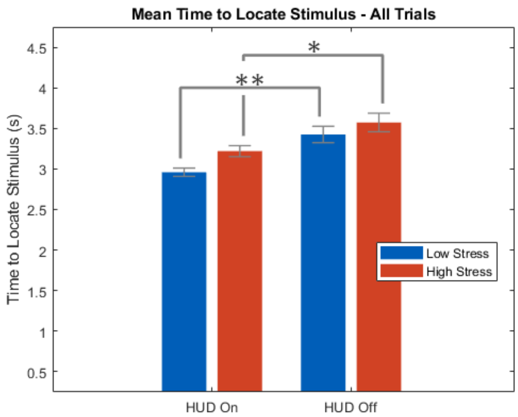


Figure 4.8: Time taken to localize stimuli, including trials where the participant failed to notice the stimulus - Stress Experiment

guidance having a significant effect overall ($F = 21.92$, $P < 0.001$) No significant difference was found for the effects of stress in pairwise comparisons both HUD conditions, however stress was found to have an overall significance ($F(1,3101) = 5.43$, $p = 0.02$). For the mean time to locate the stimulus, the mean times for the HUD-on conditions were as follows: Low-stress/HUD-on = 2.95 ± 0.05, High-stress/HUD-on = 3.22 ± 0.07. For the HUD-off conditions, the mean time to locate were as follows: Low-stress/HUD-off = 3.42 ± 0.10, High-stress/HUD-off = 3.57 ± 0.11.

For trials where participants successfully noticed the stimulus (Figure 4.9), N-Way ANOVA indicated significance in both pairwise comparisons, as well as the overall effects of the factors. For the effect of HUD guidance, a pairwise comparison of the Low-Stress

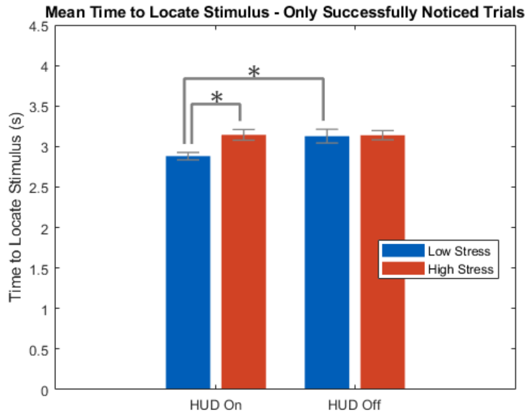


Figure 4.9: Localization time for trials where the participant successfully noticed the stimulus - Stress Experiment

conditions revealed significance ($p = 0.042$), however the effect of HUD was found to be non-significant overall ($F(1,2998) = 3.55$, $p = 0.0597$). For the effects of stress, stress was found to have an effect in the HUD-On condition ($p = 0.022$), with stress being found to have an overall effect. ($F(1,2998) = 4.52$, $p = 0.034$). For the mean time to locate the stimulus, the mean times for the HUD-on conditions were as follows: Low-stress/HUD-on = 2.88 ± 0.05, High-stress/HUD-on = 3.14 ± 0.07. For the HUD-off conditions, the mean time to locate were as follows: Low-stress/HUD-off = 3.13 ± 0.08, High-stress/HUD-off = 3.14 ± 0.06.

To further analyse participants' ability to indicate the location of stimuli, the absolute angle error of the indicated location versus the true location was investigated (Figure 4.10). During the calculation, only the Y-axis error (On a flat plane around the participant at head height) was investigated, and Z axis (height of stimuli off the ground) was not included in the calculation. N-WAY ANOVA indicated significant differences in the absolute angle error in indicated position versus true position between the HUD-on and HUD-off conditions ($P < 0.001$ for both Low- and High- Stress), with HUD guidance having an overall effect ($F(1,3101) = 556.44$, $P < 0.001$). No significant difference was found for the effects of stress in both HUD conditions. In regards to the mean values of the absolute angle difference, the mean absolute angle error was as follows: Low-stress/HUD-on = $16.57 \pm 0.81°$, High-stress/HUD-on = $19.05 \pm 0.92°$. For the HUD-off conditions, the mean absolute angle error were as follows: Low-stress/HUD-off = $50.76 \pm 1.85°$, High-stress/HUD-off = $52.57 \pm 1.83°$.

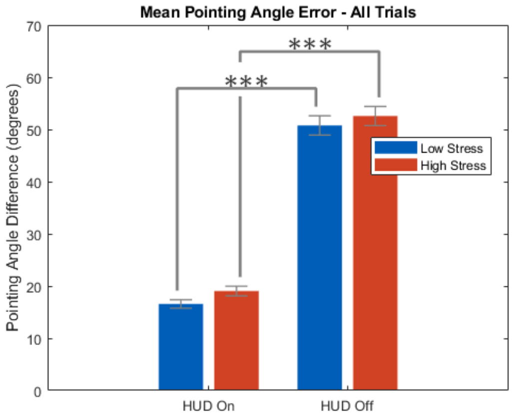


Figure 4.10: Absolute Angle Error in localizing stimuli - Stress Experiment

4.3.2.3 Identify Stage

To analyse participants' ability to identify the category of stimulus, the time taken for identification was investigated (Figure 4.11). N-way ANOVA indicated significant differences in the time taken to identify the stimuli between the HUD-on and HUD-off conditions when under stress ($P < 0.001$), with HUD guidance indicated to be a significant factor ($F(1,2998) = 22.26$, $P < 0.001$). However, in the HUD-on condition, there were no significant differences between the Low- and High- stress conditions ($p = 0.053$), with stress as a factor being found to be non-significant ($F(1,2998) = 3.53$, $p = 0.061$). For the mean time to identify the stimulus, the mean time for the HUD-on conditions were as follows: Low-stress/HUD-on = 1.89±0.04, High-stress/HUD-on = 2.05±0.07. For the HUD-off conditions, the mean time to locate were as follows: Low-stress/HUD-off = 1.75±0.03, High-stress/HUD-off = 1.76±0.03.

To further analyse participants' ability to identify the category of stimulus, the rate of successful identification (Selecting the correct category for the type of stimulus) was investigated across all trials, failed and successful. N-way ANOVA indicated no significant differences in the rate of successful identification in pairwise comparisons, as well as no significant effect of stress or HUD guidance overall. However, For successfully identifying trials where participants failed to notice the stimulus (Figure 4.12) ('failed' trials), N-Way ANOVA identified significance in the effects of HUD guidance. A pairwise comparison between HUD conditions for the low-stress state revealed significance ($p = 0.025$), with HUD guidance being a significant factor overall ($F(1,40) = 15.81$, $P < 0.001$). For the mean percentage of successful stimulus identifications after failing to notice the stimulus, in the HUD-on condition, the rate of success for low-stress was 83.33±5.77%,

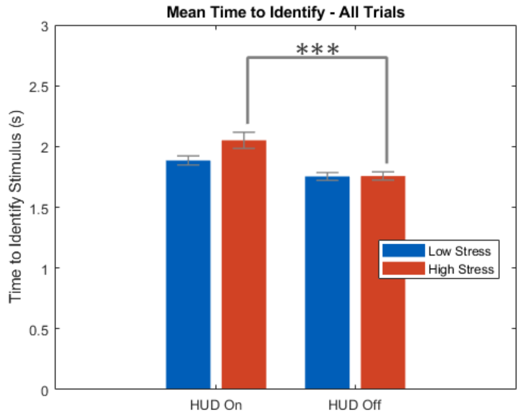


Figure 4.11: Time taken to identify the stimulus category - Stress Experiment

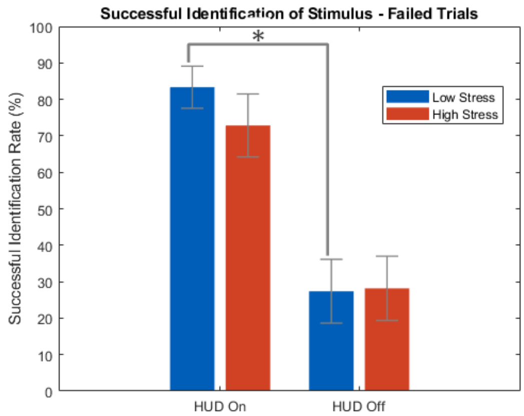


Figure 4.12: Identification Success Rate for trials where the participant failed to notice the stimulus - Stress Experiment

whereas the rate of success for high-stress was 72.86±8.63%. In the HUD-off condition, the rate of success for low-stress was 27.38±8.75%, and the rate of success for high-stress was 28.17±8.82%. A full table of the effects of each variable on task performance can be seen in table 4.2, including the interaction effects for each performance metric.

4.4 Discussion

4.4.1 HUD Performance

The results of the target localization and identification task revealed significant performance increases in success metrics for each stage of a trial, as well as significant

Metric	Effect of HUD	Effect of Stress	Interaction Effect
Notice			
Notice Time	*** ∣ P <0.001	* ∣ p = 0.017	NS ∣ p = 0.36
Notice Success	NS ∣ p = 0.075	NS ∣ p = 0.99	NS ∣ p = 0.98
Locate			
Locate Time	*** ∣ P <0.001	* ∣ p = 0.012	NS ∣ p = 0.52
Locate Error	*** ∣ P <0.001	NS ∣ p = 0.14	NS ∣ p = 0.82
Identify			
Identify Time	*** ∣ P <0.001	NS ∣ p = 0.06	NS ∣ p = 0.07
Identify Success	NS ∣ p = 0.42	NS ∣ p = 0.86	NS ∣ p = 0.97

Table 4.2: ANOVA - Effects of factor on task stage. ∣ * = P < 0.05 ∣ ** = P < 0.01 ∣ *** = P < 0.001

reductions in times taken for each stage apart from Identification, where HUD-assisted trials required a significantly larger amount of time to identify when in a state of high stress. This supports, and is supported by, previous research which has shown HUDs to provide effective assistance in comparison to an unaided condition [96, 75]. Furthermore, the HUD continued to show increased performances when participants were placed under acute stress, though did demonstrate a reduction in performance in the identification stage, potentially due to the requirement to process and recall audio-visual information as opposed to simply audio, which may prove more difficult due to the effect of stress on attention mechanisms and memory [173, 44], with stress facilitating attentional disengagement [174], potentially leading to additional cognitive work being required to recall the presented stimulus. This could further support the performance increases in the awareness and localization stages, with the response speed increase when under mild acute stress [175] being counteracted by the additional mental workload. Finally, a key aspect of the audio-visual guidance of the HUD can be seen in the identification of trials where participants failed to notice the stimulus Figure 4.12, where providing a visual indicator as to the category of stimulus significantly improved identification rates, showing a key capability of visual information delivery when there is a failure in an auditory perception capacity.

Overall, HUD guidance can be seen to continue to provide tangible benefits in aiding the noticing and localization of non-line-of-sight auditory stimuli, improving reaction times in the notice and locate stages, as well as improving accuracy during localization and during the identification of stimuli the participant failed to notice, however the effect of this additional information can be seen to be detrimental when stressed, with identification time taking significantly longer when under stress.

It is, however, essential to consider the implementation of the HUD and how it is displayed to the user, as certain design paradigms that function in wide field-of-view VR systems may not be suitable for limited field-of-view AR systems, and so considerations must be made for ensuring the continued effectiveness of these displays [64, 66, 176]. A potential area for further review is the growing industry of virtual and augmented reality applications, as with the adoption of VR into the mainstream the industry's understanding of effective XR HUD design continues to mature, and this may provide additional design guidelines for methods of implementing effective HUD designs.

4.4.1.1 Information Delivery

In dangerous environments, the ability to effectively perceive, interpret, and act on information in the environment can be essential to both the success of tasks as well as the safety of the individual performing the task, where failure can lead to catastrophic consequences. As part of this, care must be taken to ensure that any solutions designed to augment one factor must not hinder others. In situations where switching focus between the HUD and the environment is crucial, such as first-person video games, effectively switching between the environment and the HUD is a core consideration in HUD design [154, 177], leading credence in adopting aspects of their design into industrial applications where this is a critical factor [178]. Additionally, depending on the situation, the information the user actually requires, as well as the information the user perceives themselves to require (which are not always the same, or may contain excess information), may differ [179, 180, 95], and may potentially change depending on the level of stress the user is under, with users potentially preferring an excess of information when in low-stress and only the critical information when in high-stress. Having an adaptive system such as this could further support users' effectiveness in their task [181], however consideration must be taken to ensure a user-centric approach is taken.

Finally, the method of information delivery can directly contribute to the effectiveness of a HUD-based system. By providing a visual element to a stimulus that is originally solely audio-based, the methods by which the stimulus can be perceived is increased, as well as distributing the workload across various senses, providing an opportunity to notice it with one sense, should the other sense be overloaded or occupied. This multi-modal information delivery could account for the performance improvement, and the possibility of which senses are used for information delivery, and which may be negatively impacted by the environment, should be considered when designing HUD-based systems. Furthermore, as noted in prior literature [182, 183], in mental processing, the visuospatial sketchpad (Visual) and phonological loop (Audio) processing capabilities in the brain are often considered separate, and do not use overlapping resources during stimulus processing and retention in memory, therefore this distribution to multiple senses can also be seen as a potential factor in the HUD's effectiveness.

4.4.2 Effects of Stress

The effects of the acute stress invoked during the experiment can be seen in the Localization stage of the trial, where the audio-visual HUD condition took significantly longer for participants to locate the stimulus when in a stressed state, as well as ANOVA revealing an effect of stress in the time taken to complete the Notice and Locate stages of a trial. This is supported by previous literature which suggests stress has an adverse effect on auditory and visual processing capability [71, 184]. However, this study did not find adverse effects on other behavioural measurements, which could possibly be due to mild acute stress being found to also improve performance in certain tasks [175, 185], as well as increased arousal increasing auditory detection capabilities [186]. This is further supported by the Yerkes-Dodson law, which finds performance increases to a point when under mental or physiological arousal, however proceeds to decrease once that threshold is exceeded [187]. Finally, the type of task being performed can affect the impact of stress on said task [188], which could provide reasoning for the reduced impact of stress on task performance.

Additionally, participants' reported emotional state on the Self Assessment Manikin showed higher levels of arousal, indicating a higher state of excitement, as well as a lower level of dominance, indicating a sense of lack of control of the situation. However, this significance diminished towards the latter half of the experiment, becoming non-significant for arousal by the end of the third block, and dominance by the end of the final block, potentially indicating participants became accustomed to the experiment environment. This acclimatization to the stressful environment could reduce the impact of stress as the condition continued, further mitigating the effect of stress on task performance.

4.5 Summary

This chapter presents the results of a 3D virtual reality task investigating the effects of acute stress on the effectiveness of a common HUD design for the purposes of non-line-of-sight auditory target localization and identification.

The results of this study indicate that stress has a notable impact on the time taken to notice and localize auditory stimuli, leading to a potential degradation of performance when in a state of stress. The results of this study also indicate that HUD guidance continues to function when in a state of heightened stress, continuing to provide significant benefits to target localization and identification regardless of stress state, suggesting the suitability of common HUD design elements for non-line of sight target localization and identification purposes when the user is expected to be in a state of potential stress.

4.5.1 Limitations

One of the limitations of this work that warrants further investigation is the level of stress participants underwent. Whilst it is generally expected that performance degrades under high levels of stress, this study found minimal effects of stress on behavioural performance, though still found stress to be significant in regards to participants' subjective responses. This suggests that further work could be done investigating higher stress levels, to investigate whether HUDs continue to reduce a loss in performance at higher stress levels.

Another limitation was that of the sound spatializer and head-related transfer functions (HRTFs). At the time of data recording, the spatializer used (The Vive VR 3D Spatializer) was the most appropriate 3D spatializer to use. However, recent advancements in spatialization techniques that incorporate participant-specific HRTFs have led to an increase in 3D audio simulation. Whilst a pilot study revealed that the Vive spatializer functioned appropriately and effectively, it is possible that more accurate localization results may arise from using a spatializer and HRTF directly tailored to each individual participant.

Furthermore, it is possible that the task itself proved to be of a benefit for the effectiveness of the HUD. As participants were actively scanning the scene to perform the SAGAT task, their primary focus was on their visual senses, whilst their audio senses were not the primary focus. This delivery of information to the sense that is in primary use may aide in reaction times, whereas a smaller difference may have been seen if the secondary task was audio-based instead.

As mentioned in the discussion, the provision of additional information does not necessarily always pose a benefit, as whilst improvements were seen here, should one of the visuospatial sketchpad or phonological loop be overloaded with information, this could potentially become a detriment to user performance. This does raise an interesting concern and limitation, as this augmentation of a sense which is the primarily used sense may also open the possibility of information overload, and so care should be taken when designing similar systems in order to minimise the chances of information overload.

In regards to future work, further research directions include both additional environmental factors that may affect HUD usage that could be encountered in the workplace, such as fatigue, as well as further investigation into the effect of higher levels of stress on HUD usage, to investigate the limitations of video-game inspired HUDs under higher levels of a stress factor.

CHAPTER 5

How Circadian Fatigue affects HUD use and SA for Target Localization and Identification

5.1 Overview

There are many factors that can affect one's situational awareness, with fatigue being a major factor in human performance and awareness of their environment. As certain professions require long hours over the course of multiple days, such as oil rig workers or firefighters, fatigue can become a large factor in these long-term deployments in the workplace, and therefore mitigating the effects of fatigue becomes critical in ensuring adequate performance in task completion. While studies have been done on the effects of fatigue in regards to computer use and task completion, very few have investigated the effects of fatigue on AR assistance for audio localization. As perception is a core component of SA, it is critical that individuals receive the support they need for task completion. However, as more information is provided in an audio-visual form, it must be ensured that such a support system does not overload the user or cause performance degradation.

As AR has been identified as a method of providing a visual aspect to auditory aspects of the environment, Heads-Up Displays (HUDs) have continued to be used to provide localization assistance. However, the long-term effects of fatigue on HUD usage has been primarily limited to stationary use, such as that in trucks or aircraft, where the user is not required to move within their environment, and is solely focused on the task at hand,

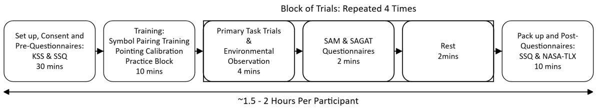


Figure 5.1: A timeline for the full Fatigue Experiment

normally the piloting of the vehicle. Therefore there is an identified need to investigate the effects of fatigue on personal HUDs in a realistic environment, where awareness of your overall surroundings may not be the primary task.

In order to investigate these effects of physiological factors on SA performance, the effects of circadian fatigue (sleep deprivation) was investigated. Additionally, the previously used AR support system was also investigated for suitability in improve performance when in a state of circadian fatigue.

5.2 Methodology

To understand the impact of circadian fatigue on sound source identification and localization using HUD, we conducted an extensive longitudinal study spanning 2-4 months. An overview of the experimental procedure is depicted in Figure 5.1. Initially, 20 participants were recruited, who agreed to wear a smartwatch for continuous sleep data monitoring, provide daily sleep condition reports, and perform a daily psychomotor vigilance task. One participant was discounted due to non-compliance with longitudinal protocol, leaving 19 participants for analysis (N = 19). These participants were called back to the laboratory for a VR experiment twice ‚Äì once during a state of low fatigue and once during high fatigue, as determined by a combination of a prediction of the participant's fatigue state supported with participant's reported fatigue state on the day. Upon arriving at the lab, participants engaged in a sound source identification and localization VR experiment, to investigate the effects of fatigue on their ability to notice, locate, and identify auditory stimuli with and without HUD guidance. The experiment procedure is explained in more detail in the following sections.

5.2.1 Participants

19 participants were recruited from the local university population. All participants were right-handed and were aged between 18-35 (Mean age = 24.35±5.85), 5 of which were female. All participants were healthy and had normal or corrected-to-normal visual

and auditory function. No participants reported any history of psychiatric disorders, neurological diseases, sleep or fatigue-related disorders, or alcohol or drug abuse. In addition, participants were required to follow protocols regarding both sleep duration and drug intake, with participants required to maintain a minimum of 6 hours continuous sleep per night as much as possible over the course of the study, as well as to consume no more than 3 standard cups of coffee or other stimulants (200mg caffeine) in the 24 hours preceding a recording session, and no caffeine in the 10 hours preceding a recording session. Participants were also required to inform the research team through a sleep diary if they took sleep aid medications, or began long-term pharmaceutical treatment.

Each participant received 80 AUD total for their participation in the virtual reality experiment sessions, and was awarded the smartwatch they wore during the longitudinal aspect as compensation. Before the experiment, informed written consent was obtained from all participants. The Institutional Human Research Ethics Committee approved this study protocol involving human subjects (ID Anonymised).

5.2.2 Circadian Fatigue Recording

5.2.2.1 Circadian Fatigue State Monitoring

To determine circadian fatigue states, participants were required to wear a Fitbit Sense smartwatch to track their nightly sleeping patterns for a minimum of two months, with the recording period ranging between 2-4 months, depending on the participant's suitability for the high and low fatigue recording sessions during the duration of the longitudinal aspect of the study as the variance in circadian rhythms of participants led to a limited number of suitable recording sessions (Mean longitudinal recording time 92$\pm$27 days). To ensure participants were correctly completing the requirements of this study, an automated system alerted the researchers to anomalous or missing sleep data, and participants received an automated text message informing them to verify with the research team the reason for the missing or anomalous data. In the event of continuous missing data, participants were contacted by the research team to ensure timely completion of the study.

The sleep data was then used with the Fatigue Impairment Prediction Suite (FIPS) [93] to predict their fatigue levels for the following week to identify potential recording sessions for a high- or low- fatigue state. To do this, sleep duration data collected from the participants' smartwatches was input into FIPS' biomathematical model, which provided a prediction of the following week's fatigue levels, in the form of predicted Karolinska

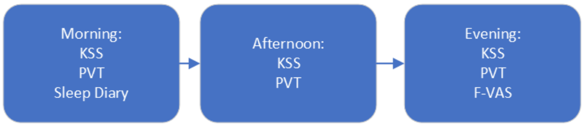


Figure 5.2: Daily routine for surveys delivered through participants' phones

Sleepiness Score values. In order to ensure FIPS' accuracy, two weeks of verification data was collected before participants were eligible to be called in for recording sessions. The FIPS predictions then identified possible recording slots, which were verified through a morning KSS survey before calling participants in for the recording session. The research team aimed to ensure a minimum of two weeks between recording sessions to limit the effects of learning on the results of the experiment, treating the time between sessions as both the study gap and the test delay [189], whilst still ensuring timely completion of the study. This resulted in an average gap between sessions of 24±4 days. In addition, efforts were made to counterbalance the order in which fatigue conditions were performed, with half the cohort performing the High Fatigue condition first, and half performing the Low Fatigue condition first.

5.2.2.2 Perceived Fatigue Level

In order to evaluate the participants' perceived fatigue levels, that is the level of fatigue which participants self-report, participants were required to install a fatigue level tracking phone app, developed by the research team for this study. The app provided data to support the reported fatigue levels through three questionnaires and one vigilance task which the participants were required to complete multiple times per day, in the morning, afternoon, and evening, on a daily basis (Figure 5.2). The questionnaires used were the Karolinska Sleepiness Scale (KSS) questionnaire [190, 191], the Visual Analogue Scale to Evaluate Fatigue Severity (VAS-F) [192], and a sleep diary delivered in the morning to report any unusual changes. A Pittsburgh Sleep Quality Index [193] survey was also delivered monthly in order to evaluate the previous month's sleep quality. Participants were required to answer the KSS three times a day, once in the morning, afternoon, and the evening. The VAS-F was delivered during the evening surveys. To measure reaction times, a visual psychomotor vigilance task (PVT) was built into the phone app to serve as a vigilance task to evaluate their psychomotor performance, which is an indicator of their fatigue level [194, 195], which participants were required to complete during each of the three survey windows.

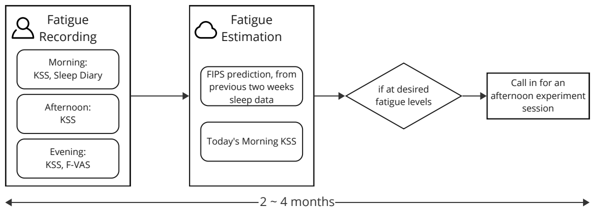


Figure 5.3: Longitudinal call-in process for recording sessions

5.2.2.3 Circadian Fatigue State Estimation

To accurately identify participant fatigue states for recording sessions, a twofold approach was taken. Firstly, data on sleep duration collected from the watch for the previous two weeks was input into the Fatigue Impairment Prediction Suite (FIPS) in order to generate a prediction as to fatigue states in the following week as a baseline (Figure 5.3). Secondly, the participants' provided KSS responses were cross-validated with the FIPS-predicted KSS scores in order to ensure a high likelihood of a suitable fatigue state during the recording sessions, which were primarily scheduled in the afternoons as these were identified as the primary times of peak fatigue.

The predictions from FIPS were generated using the inbuilt Three-Process Model (TPM), a biomathematical model based on works from Ingre et al [196]. This model was selected over the alternative 'Unified' model due to the TPM providing predictions of overall fatigue and alertness levels from the KSS, whereas the inbuilt Unified model is specifically designed to predict lapses in PVT task completion [93].

In the event that FIPS predicted a state of high fatigue during a day, and the participant's reported KSS for that morning matched or was higher than the prediction, the participant was called into a recording session on the same day. Upon arrival to the recording session, participants were required to complete another KSS as to their current state, which was compared to the prediction and noted as the KSS level for that recording session if their reported KSS matched or exceeded the predicted KSS (Figure 5.3).

5.2.3 Sound Source Identification and Localization Experiment

In order to measure participants' ability to accurately notice, locate, and identify auditory stimuli when in a state of fatigue, participants were required to complete two experiment sessions involving an audio localization task, once in a high fatigue state, and once in a low fatigue state. To counterbalance any learning effect, half the cohort completed the high fatigue session first, while the other half completed the low fatigue block first. The length of this experiment was approximately two hours.

5.2.3.1 Apparatus

In the recording session, participants wore a Vive Pro Eye virtual reality headset (2 x 1440 x 1600 resolution, 90Hz refresh rate, 3D spatial sound, and 120 Hz eye tracker sampling rate), with pupil diameter and eye tracking calibrated before each recording session. To further ensure the auditory stimuli were spatialized correctly, the official Vive 3D Sound Spatializer Plugin (version 1.2.0) [197] was used, as well, as a pilot study was conducted to verify the spatializer was functioning correctly. The spatializer used the default Head Related Transfer Function provided by the spatialzer plugin. Actions taken in the environment, such as locating the stimuli or selecting the category were done through the Vive Pro Eye controllers, with the left controller for the 'notice' stage, and the right controller for all other actions.

The specific design for the heads-up-display as well as the symbology used was based off a usability study on HUD designs by Tian & Minton et al.[159], where the chosen HUD design (Radar+Indicator) was identified as displaying similar performance to HUD designs employed in industrial applications, as well as being identified by users as preferred in several categories, including readability, emphasis on information, and interpretability [159]. This preference by users that may not have HUD experience, ability to quickly gain proficiency even with little prior use of HUDs, ease of use, and emphasis on relevant information made it suitable for a HUD design for use in a fatigue state, as the simplicity and ease of use was hypothesized to reduce the likelihood of attentional tunnelling or information overload occurring when in a state of fatigue.

The same categories of stimuli and associated symbols were used as in the previously detailed usability study to ensure similarity to the previous implementation of the HUD (Figure 5.4).

5.2.3.2 Conditions

This study used a within-subject experiment design with two independent variables, *Fatigue Level* and *HUD Guidance*. Each variable has two levels, namely (High Fatigue, Low Fatigue) and (HUD On, HUD Off). The variation of the fatigue level allows the investigation of the effect of fatigue on non-visual auditory localization, and the variation of visual guidance through the form of the HUD allows both the effectiveness of visual guidance to be seen and, most importantly, the effect of fatigue on HUD usage. To

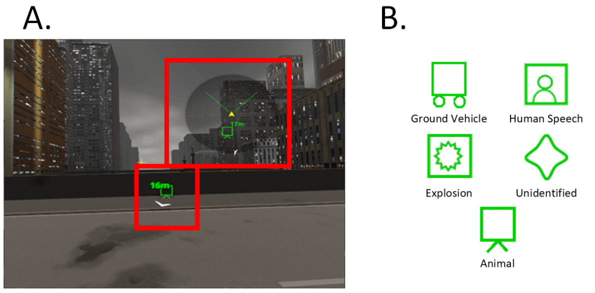


Figure 5.4: A) HUD elements used to indicated stimulus. B) All possible symbols that could appear on the HUD.

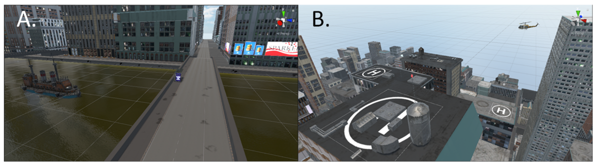


Figure 5.5: The two scenes used during the Fatigue experiment

counterbalance fatigue, half the participant cohort was required to complete the high fatigue condition before the low fatigue condition, and vice versa for the other half of the cohort. To account for learning during the virtual reality experiment, guided (HUD-On) and unguided (HUD-Off) trials were intermixed so that participants were not able to predict whether a trial was going to be guided or unguided, as well as to prevent learning or reliance on the HUD for prolonged periods of time. Additionally, two separate scenes were used for each condition, as to prevent learning of the environment between recording sessions - These scenes were counterbalanced, with half of the participants completing the High Fatigue session in Scene 1 and the Low Fatigue session in Scene 2, and half the participants completing the sessions in the opposite scenes (Figure 5.5+.

5.2.3.3 Scenario

The experiment involved participants completing two tasks during the experiment, with the primary task being to respond to, locate, and identify non-visual ('non-line-of-sight') audio stimuli within the environment, with and without visual guidance through a heads-up display. This primary task was based on previous 3D audio localization experiments [198, 199], including the addition of a virtual environment to aid localization [155].

In the inter-trial period, participants had to perform the secondary task, which was to survey the surroundings in the virtual environment in preparation for questions about the contents and state of the virtual environment, in the style of SAGAT [11], with participants being asked questions regarding the environment in the rest periods between trial blocks. This secondary task was chosen to both emulate the requirement of maintaining awareness of the natural environment as well as relying on the HUD, in addition to attempting to ensure participants' attention was not entirely on the HUD, in an attempt to mitigate the attentional tunneling effect that HUDs have been displayed to invoke [200, 201]. As participants received no warning when a trial was about to begin, they had to maintain awareness of their auditory surroundings and interrupt the secondary task when they believed a trial had commenced, simulating the requirement in natural use of the HUD in switching between integrating information from the HUD, and perceiving the natural environment.

In order to ensure participants fully understood the task, participants received three variants of training for the target localization task. Firstly, participants matched the possible sounds that could play to the appropriate symbols, training them in both the possible stimulus sounds, as well as the symbols that could appear on the HUD. Secondly, participants engaged in a calibration task where visible stimuli spheres appeared in a clockwise rotation around them, to train them in the controls for indicating the direction of a stimulus. Finally, participants completed a training block of 40 trials & post-block questionnaires before beginning the main experiment, in order to provide training in the sound localization [202].

During the experiment, participants were required to stand in a stationary spot while wearing the VR headset, while presented with a realistic virtual environment. At random intervals (5-7s after the end of a previous trial), participants were required to complete a stimulus trial. Stimuli were presented in one of the four cardinal directions (North, South, East, West), modified by up to 20 degrees in order to prevent muscle memory, such that participants required full awareness of their environment. Each of the 5 categories was presented 4 times (2 with HUD guidance, 2 without HUD guidance) in each of the 4 directions, totalling 80 trials. These trials were intermixed in direction, category, as well as whether they were HUD guided, in order to prevent learning effects.

In between trials, participants were required to observe and take note of the environment they were placed in, in order to answer questions about their environment that were delivered during the inter-block breaks.

In the rest break, participants were asked to answer questions regarding the environment that was in the style of SAGAT [11] in order to measure their awareness of their current environment, along with identifying their current emotional state on a

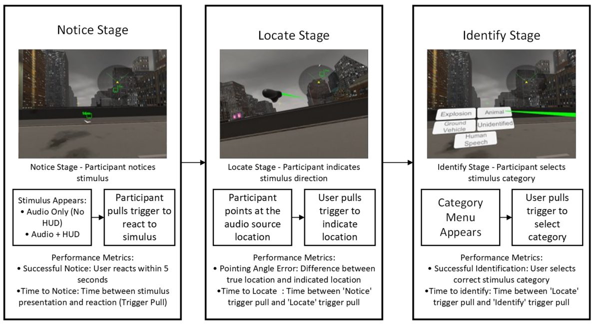


Figure 5.6: A detailed breakdown of the full procedure of one trial, including metrics and measurements

Self-Assessment Manikin (SAM) [169], as emotion has been shown to have an impact on decision making [203, 204]. In order to do this, an image of the self-assessment manikin was displayed in the scene, with participants verbally identifying where they fell on the SAM scales. After this, the screen was faded to black to prevent further inspection of the environment, and the participants were asked the SAGAT questions related to that block. All questions were given a fixed wording and were asked in a specific order.

The SAGAT questions were worded and ordered in such a way that Perception (Level 1) questions regarding a certain aspect of the environment were asked before questions regarding higher levels such as Comprehension (Level 2) or Projection (Level 3)[11, 4], so that participants were aware of that aspect of the environment, and had at least one block to observe that item in the environment, giving them time to identify that aspect of the environment before answering more complex questions regarding it. This ensured that participants were aware of an aspect of the environment in order to reduce the likelihood of non-answers on the higher-level SAGAT questions. Additionally, in order to prevent memories of previous sessions affecting SAGAT answers, two separate virtual environments were provided for the two recording sessions, with the environments being counterbalanced between fatigue levels, such that half of the participant cohort had Environment #1 for High Fatigue, and Environment #2 for Low Fatigue, and half the cohort had the opposite.

5.2.3.4 Trial Procedure:

Each trial consisted of three stages: Notice, Locate, & Identify, as seen in Figure 5.6. Participants had to complete these three stages in order, with each stage immediately following the completion of the last stage until the full completion of the trial. In the event that the participant failed the Notice Stage (Failed to indicate awareness of a trial commencing) a visual message was displayed in the centre of their vision, instructing them to try to complete the proceeding stages to the best of their ability. A guidance cross was provided in the middle of the participant's vision that was colour-coded according to the current stage of the trial, to aid participants in identifying the current stage of the trial, as well as the next actions they had to take.

In the Notice Stage, participants were required to press the trigger on the left-handed controller when they believed a stimulus had been presented. If they were correct, the guidance cross changed colours to indicate progression to the locate stage. If the participant did not successfully indicate awareness of the stimulus within 5s of presentation, the trial automatically progressed to the Locate stage, informing the participant that they had missed an auditory stimulus, and to try to locate and identify where they believed the stimulus may have been.

In the Locate Stage, participants were required to use the right-handed controller to indicate the location of the stimulus by 'pointing' at their perceived location, then pressing the trigger on the controller, which gave feedback by turning a guidance laser projecting from the controller green. Once they had indicated the location they believed the stimulus was in, they were required to face the starting position and proceed to the Identify Stage.

In the Identify Stage, participants were required to return to the starting position and select the category of sound they believed the stimulus represented from a menu that was presented in the virtual environment. During training, the appropriate symbol was displayed next to the option buttons, however during actual trials, no visual aid was provided for symbol recognition. Once participants had selected their classification, they returned to their environmental observation task while maintaining awareness for the next stimulus presentation.

5.2.3.5 Measurements

In order to measure participant performance, two measurements can be made from each stage: The success metric, and the time taken to perform each stage. For the Notice Stage, it is possible to identify trials where participants failed to indicate awareness, and thus gain a percentage of trials that were successfully noticed, along with the time taken to press the button to identify a trial, and thus gain a time taken to react to stimulus presentation. For the Locate Stage, we can identify a success metric in the deviation of the indicated location from the true stimulus location, while we can also gain a time metric through the time taken from indicating awareness to pointing to a location and pressing the button. Finally, for the Identify Stage, we can identify a success metric in whether the participants select the correct category for the stimulus, and a reaction time metric in how long it takes them to return to the selection menu and select their choice.

5.2.4 Data Analysis

In this study, 20 participants were initially recruited, however, one participant was excluded due to irregular data from non-compliance with the longitudinal protocol. The data from the remaining 19 participants was analyzed, all of whom fully completed the longitudinal aspect as well as both virtual reality experiments. Based on G*Power [171], to achieve a power of 0.90 with an alpha level of 0.05, 19 participants should result in an anticipated effect size of 0.39 with a critical F value of 4.45.

The data relating to the longitudinal aspect of this study was analyzed in MATLAB, with simple linear regression analysis performed by calculating the Pearson correlation coefficient using the MATLAB Econometrics Toolbox, in order to identify a relationship between two variables. To determine significance, a two-tailed t-test was applied, with significance level α set to 0.05. In order to account for repeated measures of participants and the varying lengths of the longitudinal aspect of the study, all correlation plots took the first 45 days of each participant's data after the FIPS calibration period of 14 days (Days 15-60 of the collected longitudinal data).

The VR experiment data was processed in MATLAB v2020b using the MATLAB Statistics Toolbox, and was separated into four conditions, Low Fatigue HUD On, High Fatigue HUD On, Low Fatigue HUD Off, and High Fatigue HUD Off. Any outliers of more than 3 times the SD were removed before analysis of each stage. A Shapiro-Wilk Test was applied to test the normality of distribution, then Repeated Measures ANOVA was applied, with significance level α set to 0.05. In order to correct for α inflation, a Bonferroni correction was applied in post. For behavioural data, the statistical tests compared the four conditions in a pairwise manner to determine significance. In all figures, * = $p<0.05$, ** = $p<0.01$, and *** = $p<0.001$.

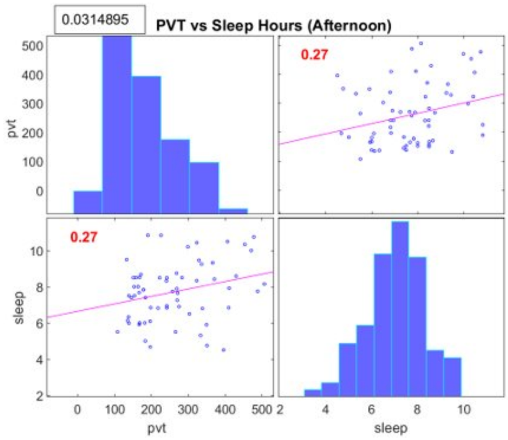


Figure 5.7: Psychomotor Vigilance Task reaction time correlated with hours slept the previous night

5.3 Results

5.3.1 Longitudinal Fatigue Measurement Results

5.3.1.1 Psychomotor Vigilance Task:

Participants were required to complete a psychomotor vigilance task (PVT) three times a day in order to evaluate their reaction speed as a metric of fatigue. This consisted of 50 trials per session, evaluated against the hours slept, the morning reported KSS, and the FIPS predicted KSS.

For the correlation between hours slept the previous night and PVT reaction time, in all sessions during the day, other than the afternoon, as well as the daily average, sleep hours were found not to have a correlation with PVT reaction time. For the afternoon session, there was a correlation with longer sleep periods showing an increased reaction time ($p = 0.03$, $R = 0.27$) (Figure 5.7).

For the relationship between KSS rating and PVT reaction time, PVT showed a moderate positive correlation as shown from the Pearson Correlation Coefficient, with the correlation being between higher perceived fatigue levels and longer reaction times. This indicates that participants' self-rated fatigue levels were an accurate supporting measure of their current fatigue state when compared to their performance on the PVT (Figure 5.8), as shown by significance in an associated t-test. ($p < 0.001$; $R = 0.32$)

For FIPS, significant correlation was not found between the FIPS-predicted KSS and PVT reaction times over the daily average, nor during any specific session. This

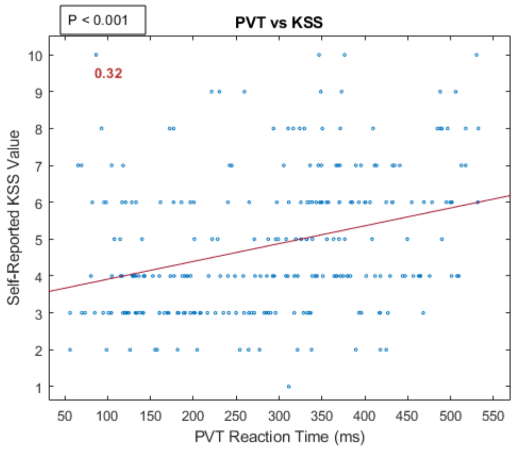


Figure 5.8: Psychomotor Vigilance Task reaction time correlated with participant-reported KSS value

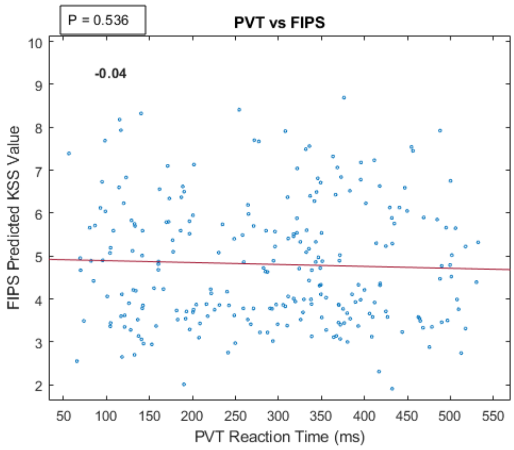


Figure 5.9: Relationship between Psychomotor Vigilance Task reaction time and FIPS predicted KSS value

suggests that using FIPS alone as an indicator of fatigue may not be appropriate, and other metrics should be integrated into fatigue state classification systems. (Figure 5.9) ($p = 0.536$; $R = -0.04$)

KSS versus FIPS:

In regards to the accuracy of FIPS, FIPS-predicted KSS and participant-reported KSS had a significant weak correlation, with FIPS predictions of a higher fatigue level generally being matched by a participant's reported KSS being an above-average value. However, these predictions were not always the same value and instead fell within a range of higher or lower values, leading to the use of the participant-rated KSS as an additional metric during the stage of inviting participants in for recording sessions.

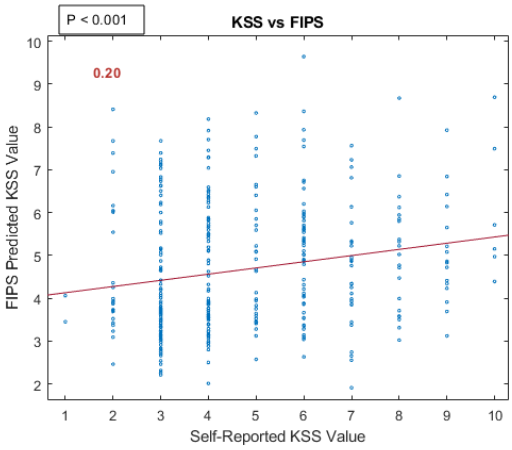


Figure 5.10: Correlation between FIPS-predicted KSS values, and participant-reported KSS values for the same time

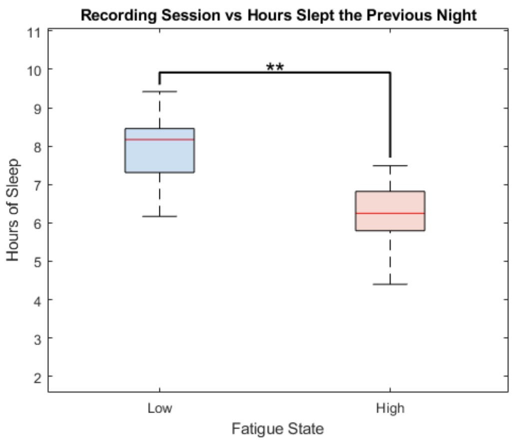


Figure 5.11: The hours slept the night before the recording session, for each condition

(Figure 5.10) ($P < 0.001$; $R = 0.20$)

5.3.1.2 Recording Conditions:

In order to determine significant differences in the fatigue states participants felt during the VR experiment sessions, sleep factors were analysed with respect to the high and low fatigue conditions. This was done to ensure the significant differences between the two fatigue conditions led to two distinct recording sessions.

For hours slept, there were significant differences between the hours slept, with the high fatigue condition having significantly less sleep the night before the experiment. ($p = 0.003$) (Figure 5.11)

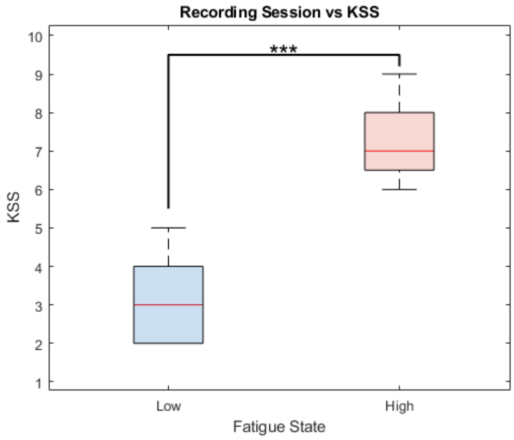


Figure 5.12: Reported KSS value for each recording condition, collected immediately before the experiment

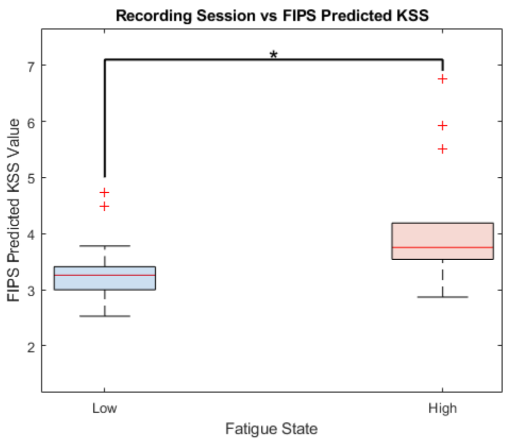


Figure 5.13: The KSS value predicted by FIPS to occur during each recording session

With respect to the participants' KSS rating which was provided before the experiment, there were significant differences between the KSSs reported for the two conditions ($P < 0.001$) (Figure 5.12)

The predicted KSS of FIPS was also investigated to ensure accurate prediction and reporting from the FIPS system. In regards to the predicted KSS during the recording session, there was a significant difference ($p = 0.02$) (Figure 5.13), however, it was significantly less than the actual reported KSS at the time of recording (Figure 5.12). This led to the consideration of the FIPS system accuracy and increased use of the KSS participants' response in the morning to invite participants into the record.

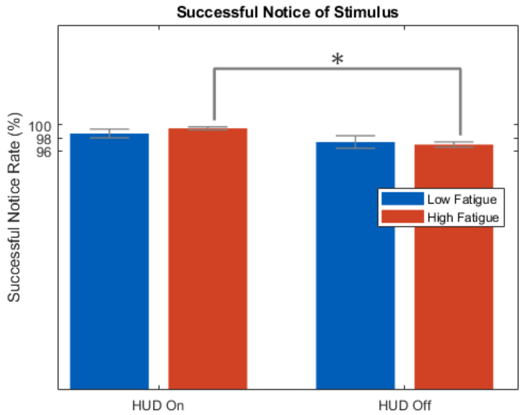


Figure 5.14: Success Rate for noticing the presented stimuli - Fatigue Experiment

5.3.2 Virtual Reality Performance Results

5.3.2.1 Notice Stage:

For the Notice Stage, there are two main metrics to evaluate: Rate of success at noticing a stimulus, and time taken to respond.

For the rate of success in noticing stimulus presentation, the results are as follows: In the Low-Fatigue HUD-On condition, the mean success rate was 98.7±0.67%. In the High-Fatigue HUD-On condition, the mean success rate was 99.5±0.24%. In the Low-Fatigue HUD-Off condition, the mean success rate was 97.4±0.96%. In the High-Fatigue HUD-Off condition, the mean success rate was 97.0±0.41%.

For the rate of success in noticing stimulus presentation, significance was found in the high fatigue condition, where participants performed significantly better with HUD guidance ($p = 0.04$) (Figure 5.14). Additionally, ANOVA reveals an effect of HUD guidance on performance, with HUD guidance performing significantly better than unguided (HUD-Off) ($F(1,75) = 9.05$, $p = 0.004$).

For the time taken in noticing stimulus presentation, the results are as follows: In the Low-Fatigue HUD-On condition, the mean time taken was 1.05±0.025s. In the High-Fatigue HUD-On condition, the mean time taken was 0.95±0.021s. In the Low-Fatigue HUD-Off condition, the mean time taken was 1.19±0.028s. In the High-Fatigue HUD-Off condition, the mean time taken was 1.07±0.022s.

For the mean time taken to notice, significance can be seen in both fatigue conditions for the effects of HUD guidance ($P < 0.001$ & $p = 0.006$ for low- and high- fatigue respectively). Significance can also be seen in both HUD-On and HUD-Off conditions

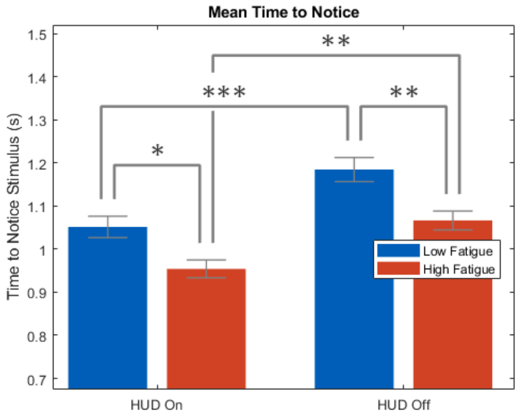


Figure 5.15: Time taken to indicate awareness of presented stimuli - Fatigue Experiment

for the effects of fatigue on task performance ($p = 0.02$ & $p = 0.003$ for HUD-On and HUD-Off respectively) (Figure 5.15). Additionally, ANOVA revealed the effects of each factor on task performance, with both HUD guidance ($F(1,2983) = 26.2$, $P < 0.001$) and fatigue ($F(1,2983) = 20.22$, $P < 0.001$) showing a significant effect on performance.

5.3.2.2 Locate Stage:

For the Locate Stage, the following metrics were investigated: Angle error in locating the stimuli, and the time taken to locate the stimuli.

For the error in localizing the stimuli, the results are as follows: In the Low-Fatigue HUD-On condition, the mean angle error was 20.38±0.87°. In the High-Fatigue HUD-On condition, the mean angle error was 17.74±0.70°. In the Low-Fatigue HUD-Off condition, the mean angle error was 38.52

±

1.53°. In the High-Fatigue HUD-Off condition, the mean angle error was 41.80

±

1.68°.

For localization error, significance can be seen in both fatigue conditions in regards to localization accuracy, with both high and low fatigue conditions having significantly improved accuracy when HUD guidance was provided ($P < 0.001$ for both conditions)(Figures 5.16-5.18). Additionally, ANOVA reveals significance in the effect of HUD guidance ($F(1,3041) = 278$, $P < 0.001$), as well as interaction effects between HUD guidance and fatigue ($F(1,3041) = 5.46$, $p = 0.02$).

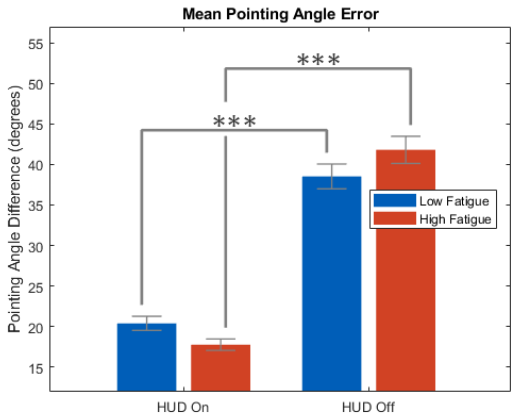


Figure 5.16: Average localization error for all stimuli - Fatigue Experiment

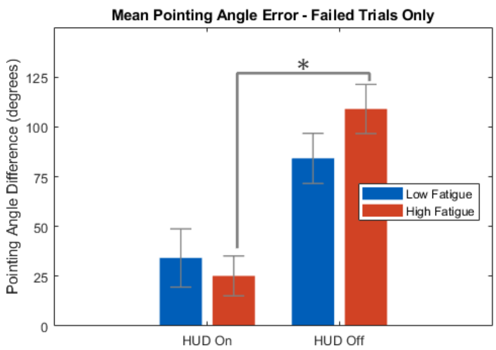


Figure 5.17: Average localization error for stimuli the participant failed to notice - Fatigue Experiment

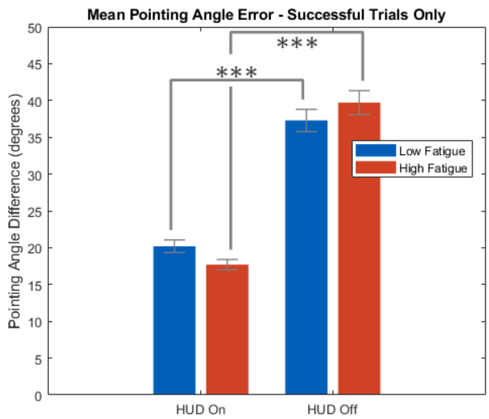


Figure 5.18: Average localization error for stimuli the participant successfully noticed - Fatigue Experiment

For the time taken in localizing the stimuli, the results are as follows: In the Low-Fatigue HUD-On condition, the mean time taken was 2.73±0.058s. In the High-Fatigue HUD-On condition, the mean time taken was 2.51±0.051s. In the Low-Fatigue HUD-Off condition, the mean time taken was 3.21±0.16s. In the High-Fatigue HUD-Off condition, the mean time taken was 2.89±0.12s.

For the time taken to locate, significance can be seen in the effects of HUD guidance, with performance improving in both guided conditions, however, significance can only be seen in the low-fatigue condition for the effects of HUD guidance ($p = 0.011$ for low-fatigue, $p = 0.07$ for high-fatigue) (Figures 5.19, 5.20). Additionally, ANOVA reveals significant effects of both HUD guidance ($F(1,3041) = 15.86$, $P < 0.001$), as well as fatigue

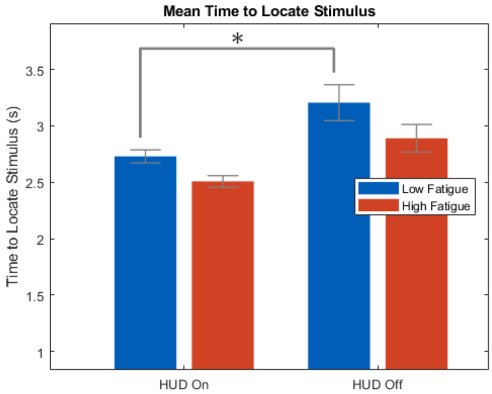


Figure 5.19: Average time to locate all stimuli, in seconds - Fatigue Experiment

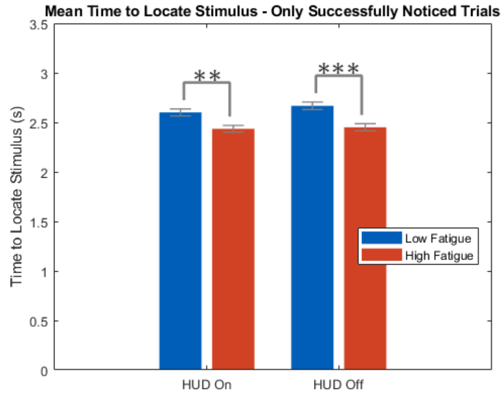


Figure 5.20: Average time to locate successfully noticed stimuli, in seconds - Fatigue Experiment

(F(1,3041) = 6.21, p = 0.013), on task performance, with task performance increasing under guidance, as well as under high fatigue.

5.3.2.3 Identify Stage:

For the Identify Stage, although no significance can be seen in comparing individual conditions in regards to correctly identifying the category of stimulus, ANOVA reveals significant effects of HUD guidance (F(1,75 = 6.64, p = 0.012) on the ability to successfully identify the stimulus category.

For the time taken to identify the stimulus, no significance can be seen when comparing individual conditions, however, ANOVA reveals effects of both HUD guidance (F(1,2983) = 7.03, p = 0.008), as well as fatigue level (F(1,2983) = 4.81, p = 0.029), on the time taken in order to identify the stimulus by selecting the category from the menu. Finally, ANOVA reveals an interaction effect in the time taken to identify stimuli, however, this interaction effect only appears in the case of trials where the participant initially failed to notice the stimulus, and required system prompting to proceed. (F(1,57) = 4.13, p = 0.047).

A full table of ANOVA P values for the effects of HUD guidance and fatigue over all three stages of a trial can be seen in Table 5.1.

Stage	Effect of Fatigue	Effect of AR	Effect of Fatigue * AR
Notice			
Notice Time	*** ∣ P <0.001	*** ∣ P <0.001	NS ∣ p = 0.664
Notice Success	NS ∣ p = 0.757	** ∣ p = 0.004	NS ∣ p = 0.354
Locate			
Locate Time	* ∣ p = 0.013	*** ∣ P <0.001	NS ∣ p = 0.663
Locate Error	NS ∣ p = 0.799	*** ∣ P <0.001	* ∣ p = 0.02
Identify			
Identify Time	* ∣ p = 0.029	** ∣ p = 0.008	NS ∣ p = 0.526
Identify Success	NS ∣ p = 0.232	* ∣ p = 0.012	NS ∣ p = 0.793

Table 5.1: ANOVA - Effects of condition on task stage for all trials. ∣ * = P < 0.05 ∣ ** = P < 0.01 ∣ *** = P < 0.001

5.3.2.4 VR Questionnaires & Environmental Awareness

For the Self Assessment Manikin (SAM), both the effects of the fatigue state, as well as the effects of the environment, were investigated. The environment was investigated to ensure there were no confounding factors from the virtual environment the participants were placed in. The only significance for the environment was found after the first block of trials, where participants showed a decreased Valence in the building condition, suggesting participants felt less happy after this block, however, this unhappiness did not persist throughout the block, with all other SAM results being non-significant in regards to the environment.

For the effects of fatigue, significance can be seen in both Valence as well as Dominance. In terms of Valence, participants felt significantly happier after the training block in the high-fatigue condition compared to the low-fatigue condition. In regards to Dominance, participants felt significantly less in control of the situation throughout the first half of the experiment in the high-fatigue condition and significantly more in control in the low-fatigue condition, with these two values eventually converging into non-significance from the end of the third block onwards. (Figure 5.21) For the Situational Awareness Global Assessment Technique (SAGAT) results, participants showed a higher level of accuracy for questions relating to perception and comprehension stages of situational awareness, however performed worse in the projection stage. This, along with the behavioural results, suggests an increase in capabilities relating to perception and basic comprehension, however high-level comprehension skills like predicting the change in environment degrade when in a state of high fatigue. (Figure 5.22)

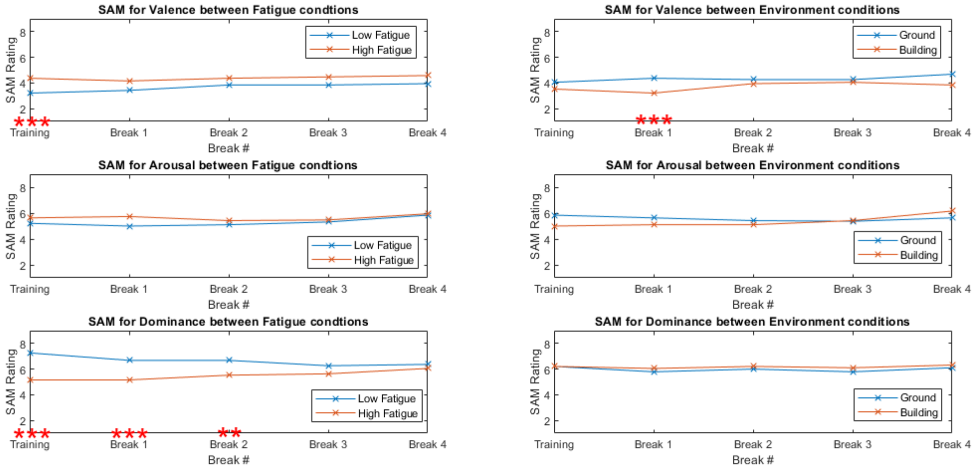


Figure 5.21: Average response on the Self Assessment Manikin during inter-block breaks - Fatigue Study

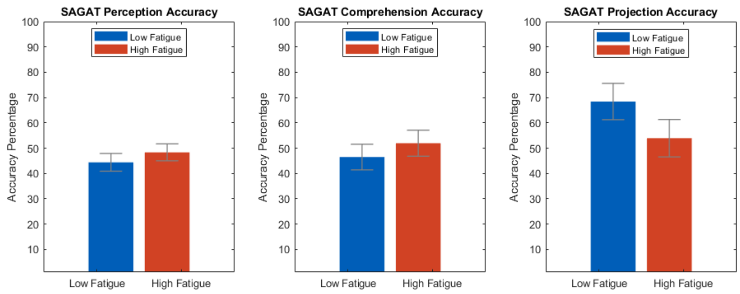


Figure 5.22: Participant score in the SAGAT section, as a percentage - Fatigue Study

5.4 Discussion

5.4.1 Predicting Circadian Fatigue with PVT and FIPS

The psychomotor vigilance task served as a baseline for gauging fatigue level, due to its use in previous studies investigating reaction time, awareness, and performance capability when in a state of fatigue, whilst also serving as a method of ensuring that participant reports on their perceived fatigue levels were consistent with their performance capabilities. As expected from previous studies, as participants reported higher perceived fatigue levels, their reaction time slowed in the PVT (Figure 5.8). This is consistent with the current consensus in literature, that a higher fatigue level results in lower performance on reaction speed-based tasks [205].

Another major finding from the longitudinal aspect of this study is the effectiveness of FIPS in predicting the effects of fatigue on performance capability. Whilst FIPS predictions are positively correlated, with reported KSS values, as per Figure 5.10, there was no significant correlation found between FIPS predicted KSS values and performance in the PVT, while significance was found when comparing reported KSS values and PVT performance. As FIPS has been shown to function accurately in situations that the model is designed for and the user cohort is suited to [206], this lack of correlation provides further support for the views of the authors of the FIPS package, which identify the requirements for biomathematical models (such as FIPS) requiring extension when introduced to new datasets and contexts [207], as whilst the model may prove suitable for the original context and data-set, the model may not function at peak efficiency when introduced to a new application unless significant parameter tuning is performed, as indicated by FIPS' reduced performance when applied to this study's participant cohort.

This highlights a key point in researching human-computer interaction under various human factors such as fatigue, where the use of biomathematical models for predicting fatigue should be used alongside traditional measurement metrics, as this combined use facilitates the extension of these models to fit new use-cases, as well as provides support for researchers in proving the model's accuracy when compared to traditional measurement metrics.

5.4.2 Circadian Fatigue Has Different Effects on Accuracy and RT

In general, HUDs provide benefits for both high- and low-fatigue conditions.

Interestingly, the effects of HUD guidance under differing fatigue conditions vary depending on the stage. This can be seen in a significant increase in accuracy in noticing stimulus presentation only appearing in the high-fatigue condition, however a significantly faster time to locate the stimulus can only be seen in the low-fatigue condition for HUD guidance (Figure 5.19), which could potentially be explained by the effects of fatigue changing depending on what the task being performed is [208].

This task-dependent difference is highly relevant to note for system design, as previous studies have identified that both fatigue and the wake maintenance zone can have differing impacts on task performance, depending on what the task is [208], leading to the requirement of consideration as to how fatigue will impact a user when making use of a system, as well as whether that effect will be consistent across subsystems and use-cases of that system.

Another interesting finding was the improvement of performance when in a state of high fatigue, when common logic would suggest a loss in performance would be observed instead. In regards to fatigue, performance improvement when in a fatigue state has been observed before [209, 210], with this improvement being hypothesized to be due to fatigued individuals actively being more alert due to their awareness of their own mental state, and thus putting more effort into task completion. Another potential explanation for the improved task performance when in a state of fatigue could the phenomena known as the 'Wake Maintenance Zone', which is the period in the average circadian rhythm (normally in the afternoon) where an increase in performance will be seen even when in a state of fatigue (When the average person will receive their 'second wind'). As the virtual reality tasks were primarily completed in the afternoon, the period of completion falls within the expected wake maintenance zone of the average circadian rhythm. As the wake maintenance zone has been shown to have a beneficial effect on neurobehavioural performance [211, 212], this could provide support for the improved performance in the fatigued condition, as the effect of the wake maintenance zone is most noticeable when in a state of sleep deprivation, and has a lesser effect when well rested [211, 212].

This unexpected finding of a performance increase, or a lack of performance degradation, is not entirely without support from previous literature. While the mechanism, causes, and effects all appear to be varied depending on the study that has been performed, a number of studies across multiple fields, ranging from an extended cognitive task [213], to extended operation of jet aircraft [214], have both found that performance degradation can be avoided or overcome, potentially through the application of increased attentional effort. This variation and non-standardisation in fatigue research, due to the relevance in multiple fields, has led to multiple frameworks being proposed for methods of classifying fatigue research to better aid researchers across disciplines [215, 216].

5.4.3 Circadian Fatigue and Environmental Awareness

While the results of the SAGAT show no statistical significance, there are still trends that can be noticed. When in a state of high fatigue, perception and base level comprehension performance increased, while high level comprehension of how the environment would change (Referred to as 'Projection' by the author of SAGAT) degraded, though not significantly. While the common consensus is that fatigue negatively impacts performance capabilities [217, 215], there is also evidence that fatigue can both improve the learning of sequences [218], such as the sequential nature of the virtual reality experiment, as well as there being compensatory neural mechanics in response to fatigue [219]. It is also known that base skills can be improved when in a state of pressure [187], such as that of performing a task when fatigued, however highly complex tasks that may require more cognitive load can suffer when in a state of fatigue, as the requirements for the task exceed capability. This may potentially explain the increase in performance during the virtual reality task, when a common consensus would be that task performance should degrade. Additionally, in the high fatigue state, the SAM results (Figure 5.21) show a higher level of unhappiness (Lower Valence) immediately on beginning the task, as well as a continued sense of a loss of control until the user is well into the task. While the design philosophy of ensuring systems are designed with the user state in mind is commonplace, this variance in effect highlights the requirements of identifying the effects of factors such as fatigue on each task the user is required to perform, as these factors may have varying effects on each sub-task the user is required to perform.

5.4.4 Design Considerations for HMD Use in Different Physical and Mental Conditions

Finally, considerations must be made when accounting for interaction effects between systems in use for the benefit of the user, and environmental factors that are affecting the user. As per Table 5.1, interaction effects can be seen in the localization error, where fatigue in the unaided condition increased the error, but fatigue in the HUD condition improved localization significantly. While in this situation the interaction results in a positive effect, fatigue can lead to a decrease in cognitive processing capability [88], which could potentially lead to a decrease in performance if the HUD or system requires more cognitive processing in order to extract meaningful information, as has been noted in literature on design principles for increasing awareness of an environment [7], as well as the implications of operator fatigue in interaction with a system [220].

This awareness of context-sensitivity suggests that considerations be made when designing interactive systems for differing contexts, as a more complex system may provide benefits for a user when in a state of low fatigue, but that benefit could transition into a hindrance if the user does not have the mental capability to process the information. This variance under different environmental conditions can be identified as a design point when developing systems for human-computer interaction. While it is possible for systems to be designed in such a way as to be able to function in multiple conditions with a minimal loss or gain in efficiency between conditions, this compromise on function may reduce the performance of the system overall. Another potential approach could be to identify the user's current state and tailor the user experience to optimize towards that state. This avenue of adaptive interaction based on the user's current state may, however, require further research in order to identify whether users are able to effectively adjust to a changing system when in a fatigue state, whilst research into the effectiveness of adaptive systems has been conducted [221, 181, 222], the investigation of the effect of human factors on these systems, such as circadian fatigue, has yet to be investigated.

5.5 Summary

The results of the VR sessions reveal an improvement in performance when making use of an HMD-based HUD for localization, suggesting the suitability of HUD-based guidance systems for localization and identification purposes. However, a decline in complex cognitive processes can be seen in a state of high fatigue, with both time taken to identify stimuli categories, as well as projection of the future state of the virtual environment showing reduced performance, as is expected when in a fatigued state, suggesting that other fatigue-related issues may continue to persist even with visual augmentation, and may require further work in order to mitigate the effects of fatigue. From the results of this study, this chapter suggests design considerations in regard to designing systems for human-computer interaction when in a fatigue state, as well as for systems that may be used when the user is in varying levels of fatigue.

5.6 Limitations and Future Works

A limitation identified during this study was the capability to accurately predict participant fatigue solely through the use of the FIPS package. Whilst it is possible to generate a baseline assumption, which is the case of FIPS is tailored to airline staff, there will be variance in individuals' wake maintenance zone ('second wind'), which FIPS is unable to identify due to solely being trained on sleep data, and does not account for biometric or subjective data collected during wake hours. Whilst in the study this lack of capability could be offset through the use of regular surveys during the longitudinal aspect, as well as before recording sessions to verify fatigue level, this still identifies a potential aspect of further development in fatigue prediction, as by incorporating this data, it may be possible to achieve a more accurate prediction on the individual level which accounts for variations due to individual circadian rhythms, resulting in a higher capability to account for user fatigue. Furthermore, this capability to more accurately assess fatigue levels may provide additional research capabilities, as higher accuracy in prediction as well as further temporal distance in forecasting may allow researcher to identify more potential fatigue levels in studies (Such as a 'medium' fatigue level in future works), as well as longer accurate prediction windows allowing fatigue-dependent recording sessions to be planned in advance, potentially reducing the required length of longitudinal studies that require observation for suitable recording windows.

Another fundamental limitation of this study comes in the guidance method employed through the HUD. As there exist multiple guidance methods for the localization of objects within an environment, It can be assumed that with various complexity levels and intuitiveness, different designs will perform differently when in a state of fatigue. Furthermore, the effects of fatigue were not investigated on non-visual-based guidance systems, such as that of haptic guidance. While for HUDs, previous literature suggests that the more complex and intricate the system, the more cognitive effort is required to derive the required information for localization, though the effects of fatigue on non-visual guidance systems have the possibility of further study, as it may be possible to leverage other non-visual senses to greater effect when in a state of fatigue.

5.7 Conclusion

In conclusion, this chapter presents a two-fold study into both the effectiveness of classifying user fatigue using a biomathematical model (FIPS), as well as the effects of circadian fatigue on the use of an HMD for localization and identification of auditory non-visual stimuli. The results of the longitudinal study reveal that while there is a correlation between predicted fatigue levels using only sleep duration and actual reported fatigue levels, this correlation does not necessarily carry over into actual performance in a psychomotor vigilance task. The results of the virtual reality sessions reveal an improvement in performance when making use of an HMD-based HUD for localization, as well as a general improvement in reaction times in contrast with the psychomotor vigilance task, suggesting the suitability of HUD-based guidance systems for localization and identification purposes. However, a decline in complex cognitive processes can be seen in a state of high fatigue, with both time taken to identify stimuli categories, as well as projection of the future state of the virtual environment showing reduced performance, as is expected when in a fatigued state, suggesting that other fatigue-related issues may continue to persist even with visual augmentation, and may require further work in order to mitigate the effects of fatigue. From the results of this study, the paper suggests design considerations in regard to designing systems for human-computer interaction when in a fatigue state, as well as for systems that may be used when the user is in varying levels of fatigue.

CHAPTER 6

The Brain Dynamics of Contextual Semantic Comprehension

6.1 Overview

Comprehension forms a critical step in the process of decision making and SA, with your Comprehension of a situation affecting how you Project the situation will evolve, and thus effecting the decision you eventually make. Whilst it is possible to infer user comprehension based on the action taken, being able to detect user comprehension as it occurs provides the ability to intervene at the earliest possible opportunity should an error in comprehension occur. In order to do this, previous studies have investigated the ability to detect and classify semantic attributes as they occur in the brain through EEG. To further this research, the ability to detect and classify contextual semantic attributes, that is to say semantic attributes that change depending on the situation, was investigated.

6.2 Methodology

6.2.1 Participants

20 participants were recruited from the local university population. All participants were right-handed and were aged between 18-35 (Mean age = 24.6±4.41), 7 of which were female. All participants were healthy and had normal or corrected-to-normal visual

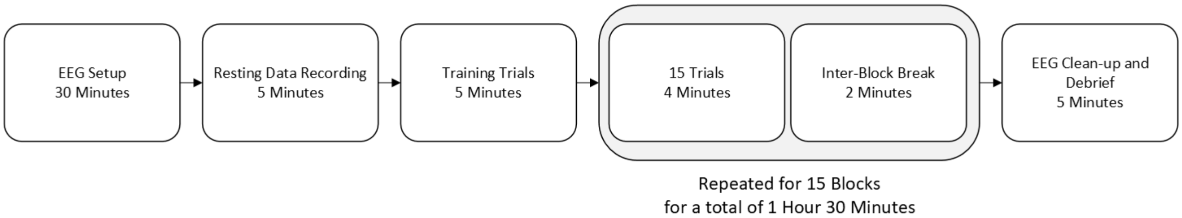


Figure 6.1: A diagram detailing the timeline for the full EEG semantic experiment

and auditory function. No participants reported any history of psychiatric disorders, neurological diseases, cognitive processing disorders, or alcohol or drug abuse. In addition, all participants had English as their first language.

During data analysis, one participant was excluded due to poor-quality EEG data, therefore the total number of participants in this study was N = 19.

6.2.2 Experiment Procedure

The complete experiment took approximately two hours, and was comprised of the following stages: EEG Setup (30 min); Resting Data Collection (5 minutes); Training (5 minutes); Task Completion (1 hour 30 minutes); EEG Clean-up and Participant Debrief (5 minutes). A visualization of the full experiment procedure can be seen in Figure 6.1.

6.2.2.1 Task Protocol

In order to limit confounding factors such as additional contextual information as well as to follow previous paradigms from similar studies, the design choice was made to prioritize simplicity within the experiment paradigm. To do this, participants were shown a context in the form of a job profession, and then shown an item. Participants then had to classify the item as either Helpful, Irrelevant, or Detrimental for if the item *was required* to be used by the profession for their usual tasks.

The experiment consisted of 225 trials, separated into 15 blocks, each consisting of 15 trials. Between each block, participants were given a 2 minute rest to prevent fatigue.

Each trial lasted 17 seconds, separated into the following parts: Focus (2s), Context (4s), Item (5s), Answer (4s), Rest (2s). Each trial was comprised of a unique pairing of Context and Item, and the participants were informed during each break that there were no repeat trials to attempt to minimize other memory effects within the EEG data. A visualization of a full trial can be seen in Figure 6.2.

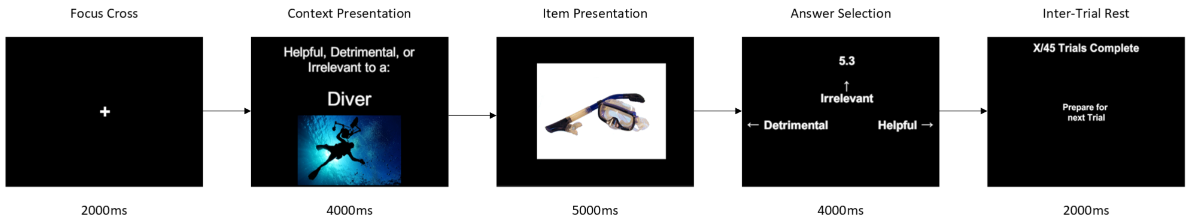


Figure 6.2: A diagram detailing the procedure of a single trial for the Semantic Comprehension experiment

6.2.2.2 Recording Equipment & EEG Protocol

Data was recorded using a 128 channel wet sensor EEG cap (Neuroscan Quik-Cap with SynAmps 2 amplifier), with channel impedance ensured below 5 kΩ before recording began. The 128 electrodes were placed according to the expanded 10-5 electrode placement, based off the 10-20 electrode placement system.

Once the EEG recording equipment had been confirmed to be below the required impedance, each participant had a baseline recorded, with 3 minutes recorded each of eyes-open, and eyes-closed resting.

In order to reduce the effect of display and input lag on features in the EEG data, the experiment was conducted on a 144Hz monitor with a mechanical keyboard as a method to reduce delays between both screen updates for images as well as reduce delay on keyboard inputs from participants.

6.2.3 Data Analysis

For this study, all EEG analysis was conducted in MATLAB v2023a, using the EEGLAB 2023.1 Toolbox. The pre-processing pipeline used for this study was based on Mokoto's Pipeline, with considerations on parameters for certain steps based on evolutions of this pipeline. Listed below in Table 6.1 are the pre-processing steps taken, including references for additional toolboxes used, such as DipFit for dipole fitting.

6.2.3.1 Components

The resulting EEG data was decomposed using EEGLAB's built-in ICA analysis tools in order to compute a number of independent components equal to the number of channels (n = 128). From here, artefact analysis was conducted using IC Labelling, with any

Step Number	Processing Step	Reference (If relevant)
1	Downsample to 250Hz	
2	Band-pass filter between 2-45Hz	
3	Clean line noise using CleanLine package	
4	Clean noise channels	
5	Interpolate missing channels from previous step	
6	Re-reference cleaned data	
7	Run ICA on cleaned data	
8	Estimate dipole location using DipFit package	
9	Epoch based on condition	

Table 6.1: The ordered steps taken to process the recorded EEG data.

component having less than 85% brain component classification being removed from component analysis before clustering.

6.2.3.2 Clustering

Clustering was performed using EEGLAB's built-in clustering feature, using K-Means clustering to produce said clusters. Clustering was performed 20 times, with a target number of 37 clusters as the optimal number of clusters. The target number was obtained through the use of the EEGLAB optimal cluster count feature, which makes use of the MATLAB Statistics and Machine Learning Toolbox's *evalclusters* function and uses a combination of a K-Means clustering algorithm, and a silhouette evaluation criterion to determine the optimal number of clusters.

After performing the clustering 20 times, The clustering set was selected that contained the most brain components in the frontal and temporal-parietal regions. Multiple-clustering (n = 20) was conducted in order to reduce random variance while still maintaining a suitable number of clusters to prevent over-fitting of the clusters.

6.3 Results

The results for this study have currently been analysed from a classification standpoint, focusing on both the initial channels for rapid acquisition, as well as ICA clusters for a more comprehensive analysis of the regional activations and post-hoc classification. Additional analysis in future analysis could also be conducted on both overall regional activation, connectivity between regions using a package such as Brainstorm, and

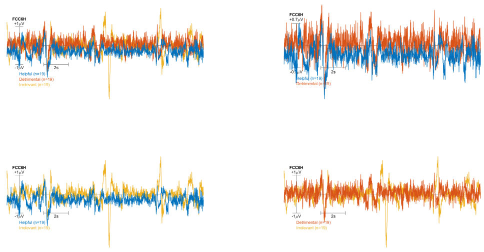


Figure 6.3: FCC6H ERP Results & Condition Comparison

information flow between region through the use of a toolbox such as SIFT to identify the process of semantic comprehension as it occurs over time.

6.3.0.1 Channels

FCC6H: FCC6H is located in the right Frontal Central region of the brain, leaning towards the Central region. In the Helpful-Detrimental comparison, Detrimental can be seen to have a larger amplitude during the classification period, as well as a potential phase shift when compared to the Helpful signal, which can also be seen to be at a negative voltage when compared to Detrimental. In the Helpful-Irrelevant comparison, Helpful can be seen to continue the trend of negativity in comparison, however a much larger response to stimulus presentation for classification can be seen in the Irrelevant condition, along with a potential phase-shift between conditions. This trend of the Irrelevant condition's distinction can also be seen in the Detrimental-Irrelevant comparison, where both a distinction at the start of the classification can be seen, as well as a potential phase-shift in the signals, however both signals fall within the same voltage range. Figure 6.3

FCC5H: FCC5H is located in the left Frontal Central region of the brain, leaning towards the Central region. In the Helpful-Detrimental comparison, Detrimental can be seen to have larger amplitudes throughout the trial, however little indication of a phase-shift can be seen, suggesting potentially limited usefulness in a classification use, however this channel may prove useful as a supplementary channel to other channels in the frontal-central region to provide an increased classification accuracy. Figure 6.4

FC6: FC6 is located in the right Frontal Central region of the brain. FC6 shows distinctions in the Helpful-Detrimental comparison, where Detrimental once again follows a trend of higher amplitudes in comparison, however no phase-shift can be seen.

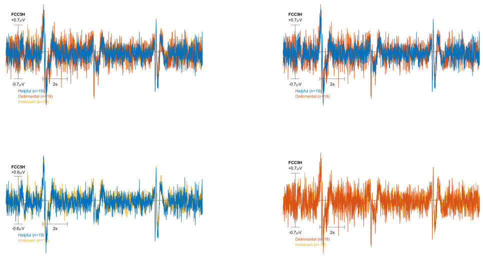


Figure 6.4: FCC5H ERP Results & Condition Comparison

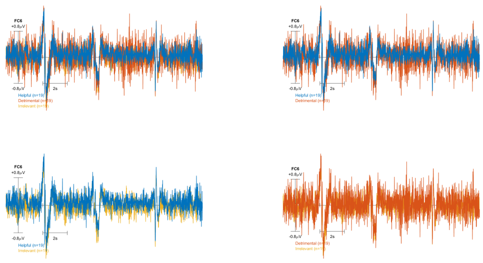


Figure 6.5: FC6 ERP Results & Condition Comparison

This increase in amplitude can also be seen in the Detrimental-Irrelevant comparison, which may provide useful distinctions for classification purposes. Figure 6.5

CPP3H: CPP3H is located in the left Central Parietal region of the brain, leaning towards the Parietal region. It shows distinctions in both the Helpful-Detrimental as well as the Helpful-Irrelevant comparisons, with Helpful-Detrimental showing more overall activity on the Detrimental ERP signal, as well as a potential phase-shift when the two signals are compared. The Helpful-Irrelevant comparison shows the Helpful condition at a larger negative voltage compared to the Irrelevant condition, with another potential phase shift when comparing the two signals. Figure 6.6

P3: P3 is located in the left Parietal region of the brain. For P3, the Helpful condition continues the trend of negativity in comparison to the other two conditions, while also showing a phase shift when compared to the Detrimental condition. This phase-shift can also be seen in the other comparisons, showing distinctions in phase between the three conditions, however no significant distinction in amplitude can be seen. Figure 6.7

PO9: PO9 is located in the left Parietal Occipital region of the brain. PO9 shows distinction in phase shift and base voltage regards. In the Helpful-Detrimental comparison, both a phase-shift can be seen, as well as Helpful having a significantly more

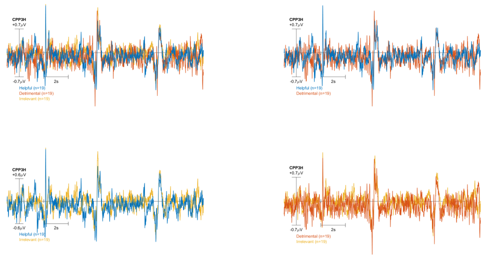


Figure 6.6: CPP3H ERP Results & Condition Comparison

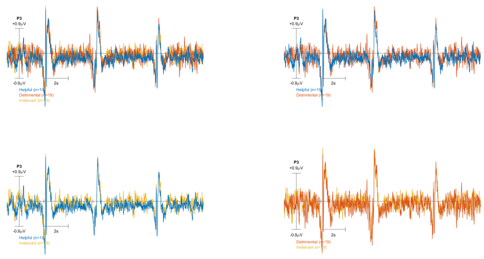


Figure 6.7: P3 ERP Results & Condition Comparison

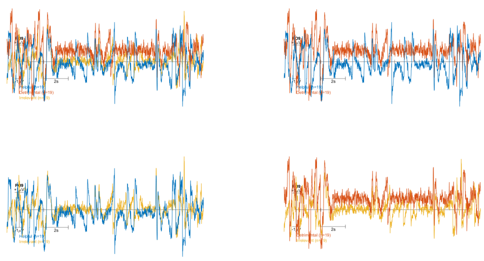


Figure 6.8: PO9 ERP Results & Condition Comparison

negative voltage than Detrimental. This phase-shift can also be seen in the Helpful-Irrelevant comparison, however the difference in voltage levels is less significant. The Detrimental-Irrelevant comparison also shows a phase-shift between the two, along with the Detrimental condition having a consistently positive level of amplitude compared to Irrelevant. Figure 6.8

POOz: POOz is the centre channel of the Parietal Occipital region of the brain, leaning towards the Occipital region. POOz continues to display the trend of phase-shifts between the conditions, with significant phase shifts seen between all three conditions.

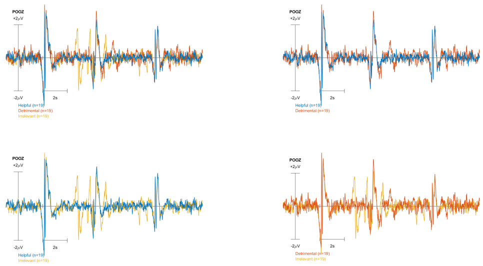


Figure 6.9: POOZ ERP Results & Condition Comparison

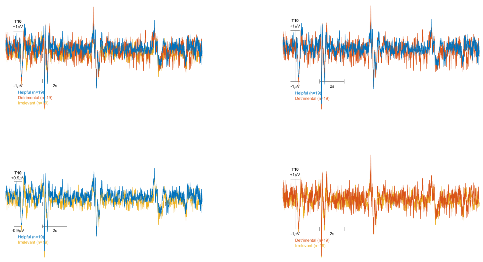


Figure 6.10: T10 ERP Results & Condition Comparison

Figure 6.9

T10: T10 in located in the right temporal region of the brain. T10 continues to display the trend of an increase of amplitude in the Helpful-Detrimental comparison, as well as a phase shift between the Irrelevant condition and the other two conditions. Figure 6.10

6.3.0.2 Clusters

After clustering, 7 clusters were identified in the areas of the brain where long-term semantic memory processes as well as cognitive decision making processes are identified to occur. As this is the area of the brain where a biomarker is most likely to occur, these are the clusters that are focused on during analysis and result presentation. Figure 6.11

Cluster 4 is located in the occipital region, primarily handling vision function. Differences can be seen in the Alpha band, primarily between 9Hz-11Hz, with both the Helpful and Detrimental condition showing a higher positive power than the Irrelevant condition. Figure 6.12

Cluster 5, located in the frontal region of the brain which handles decision making processes, shows significant negative activity throughout all bands in the Helpful and

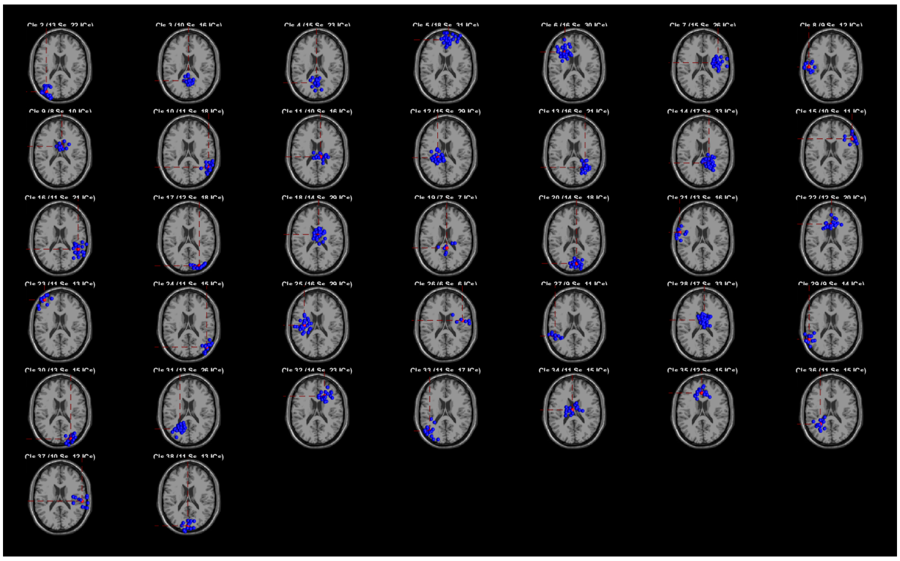


Figure 6.11: All identified EEG component clusters

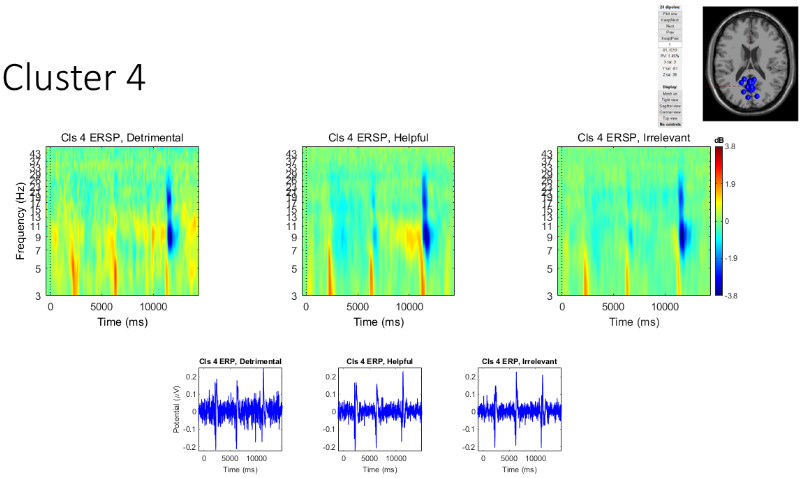


Figure 6.12: Cluster 4 Results (ERP, ERSP, Component Locations)

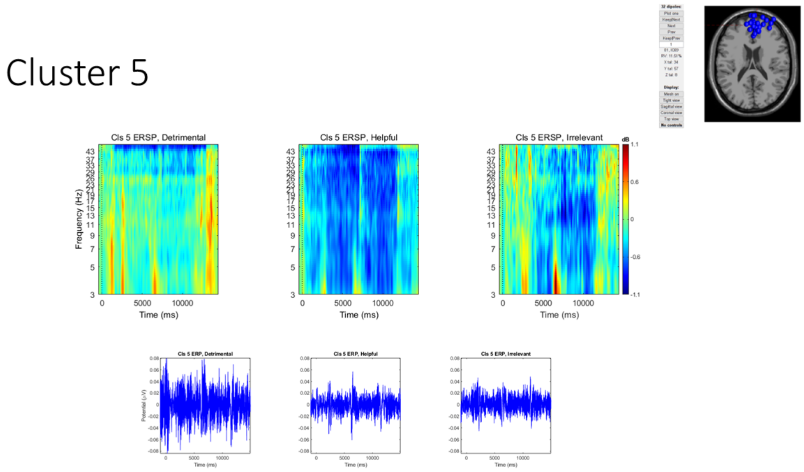


Figure 6.13: Cluster 5 Results (ERP, ERSP, Component Locations)

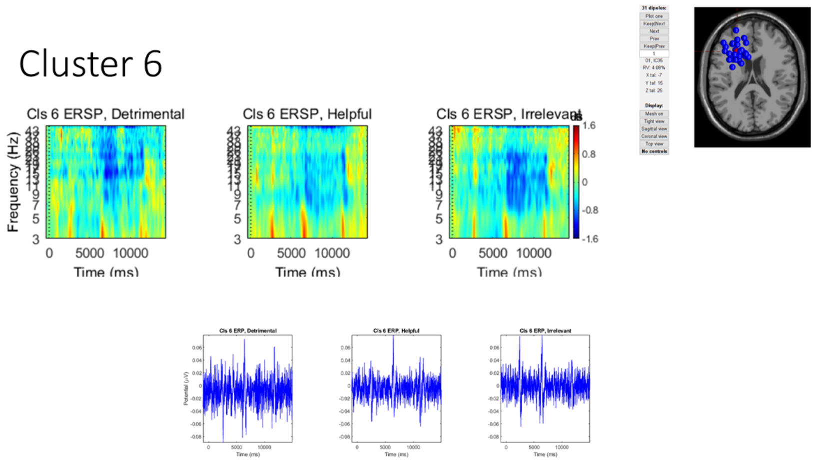


Figure 6.14: Cluster 6 Results (ERP, ERSP, Component Locations)

Irrelevant condition, however this negativity can only be seen in the Gamma band in the Detrimental condition. Figure 6.13

Cluster 6 is located in the frontal parietal region, and similarly to Cluster 5, shows large values of negativity from 10Hz-30Hz. This negativity is more pronounced in the Irrelevant condition than in the Helpful condition, with additional features in the Detrimental condition, where the negativity continues into higher frequencies than in the other two conditions. Figure 6.14

Cluster 7 is located in the right parietal region, and whilst semantic memory is shown to primarily be located in the left parietal, there are still distinctions in the condition,

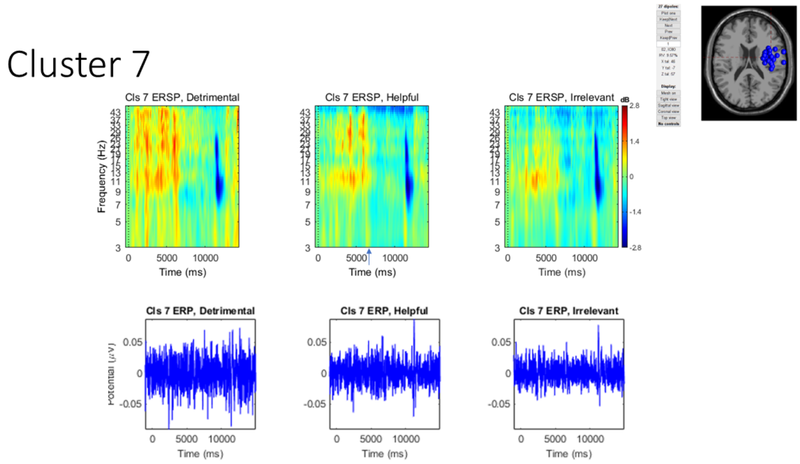


Figure 6.15: Cluster 7 Results (ERP, ERSP, Component Locations)

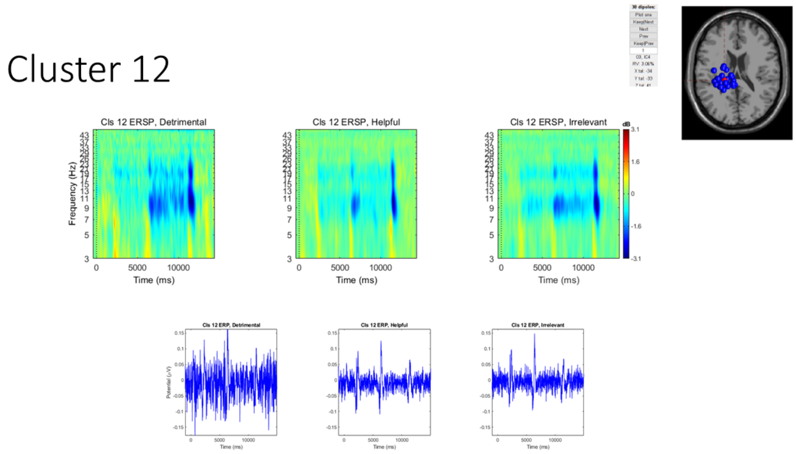


Figure 6.16: Cluster 12 Results (ERP, ERSP, Component Locations)

with the Detrimental condition showing slight tendencies towards positivity and the Irrelevant condition tending towards negativity. Figure 6.15

Cluster 12 falls within the left parietal region of the brain, where long-term semantic memory functions are indicated to take place. Distinctions can be seen in the three condition, with the Detrimental condition displaying a large period of negativity in the range of 9-25Hz, the Irrelevant condition displaying negativity in the 9-11Hz range, and the Helpful condition displaying no significant negativity through the classification period. Figure 6.16

Cluster 13 is also located in the right parietal, overlapping with the occipital region.

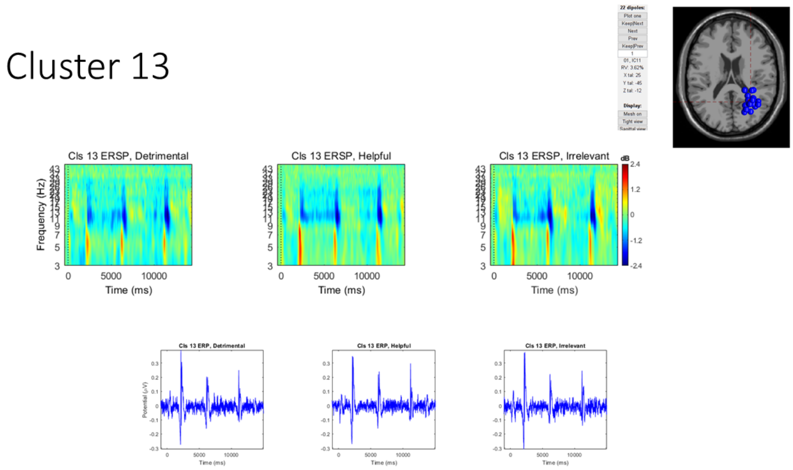


Figure 6.17: Cluster 13 Results (ERP, ERSP, Component Locations)

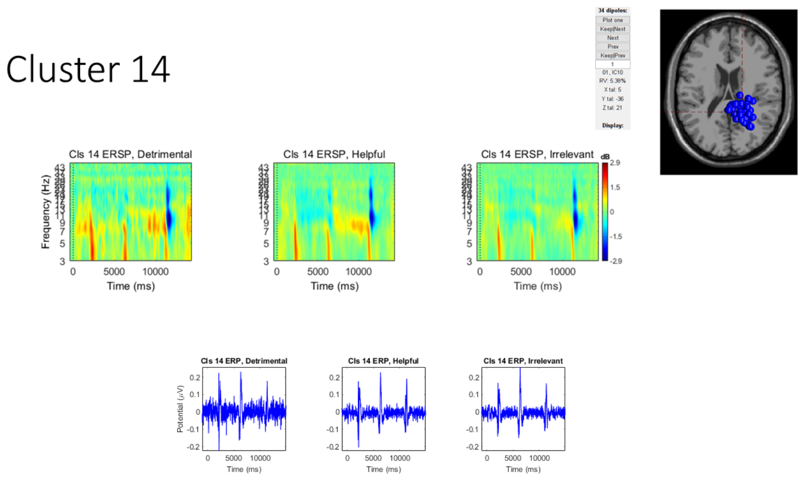


Figure 6.18: Cluster 14 Results (ERP, ERSP, Component Locations)

The trend of the Detrimental condition towards negativity can also be seen, though not as distinct as other regions. Figure 6.17

Cluster 14 is located in the central-parietal region, and continues to display significant differences between the Detrimental condition and the other conditions, with the Detrimental condition showing significant positive activation in the 9-15Hz range, the Helpful condition displaying slight positive activation in the 9-11Hz range, and the Irrelevant condition showing little to no positive activation. Figure 6.18

6.4 Summary

In summary, it can be seen that there are noticeable distinctions in both channel-based and component-based ERPs in the frontal and temporal-parietal regions, as well as distinctions in the ERSPs for components. Additionally, clustering reveals activity in the frontal region, especially in the Detrimental and Irrelevant conditions, where negative semantic relationships in the Detrimental condition and/or the lack of semantic relationship in the Irrelevant condition may lead to further consideration and processing from the participant. This finding of activity and component clusters in the frontal region is supported by previous literature, which has previously identified similar neurological processes taking place in the frontal region of the brain [140, 223]. This identification of a primary region with which to focus on data collection allows a reduction of the number of channels required, as a smaller number of channels may be used to collect EEG data from the frontal region of the brain, without extraneous sensors on unrelated areas of the scalp being required, thus increasing the portability and ease-of-use for collecting semantic data with a scalp-based EEG system.

Overall, these distinctions may allow the possibility of training a classifier to determine participant classification before the participant indicates their understanding by pressing the related button.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

7.1 Findings

In *Chapter 3*, this thesis investigated the effects of previous experience on commonly-used Heads-Up Display designs for localization. In this chapter, this thesis found that HUDs with which users had previous experience showed both improved performance on first use, as well as continued improvement over time. This can be seen in the statistical significance between the display methods with which participants were familiar (Radar-based), when contrasted with the Compass-based methods, within the experienced gamers category. Additionally, the improvement over time can be seen when the different groups and display methods are compared in Figure 3.4. Overall, this suggests that previous user experience on HUDs can be leveraged for improving SA with a HUD support system. This is especially relevant for the HCI field, as this effect of prior experience can be leveraged when designing new systems by incorporating features with which users may have previous experience.

In *Chapter 4 (RQ 1)*, this thesis expands further on the capability of commonly-used HUD designs, investigating the effects of physiological stress on situational awareness and HUD capability for improving SA. Results found in this chapter show that while certain aspects of SA in Level 1 (Perception) can improve when in a state of stress, other aspects, such as Comprehension can decrease, leading to a diminished classification accuracy. These differing effects can be most significantly seen in the significant improvement of the time taken to notice a stimulus (Figure 4.6, $p < 0.001$), when contrasted with the HUD causing identification to take significantly longer when in a state of fatigue (Figure 4.11, $p < 0.001$). Overall, this suggests that HUDs can be of benefit in augmenting situational awareness by providing additional visual cues to auditory information, however the specific use-case and intended effects must be considered before implementation.

In *Chapter 5 (RQ 2)*, this thesis continues to investigate physiological factors, with the additional investigation of Circadian Fatigue, more commonly known as sleep deprivation, on situational awareness and HUD capability, through a longitudinal study. The results from this chapter show an interesting trend, that whilst fatigue causes a decrease in accuracy normally, the addition of the HUD shows an increased accuracy of localization when in a state of fatigue when compared to a non-fatigued state, as can be seen in Figure 5.16 ($p < 0.001$). However, the increased time taken to classify stimuli ($p = 0.008$) indicates a further requirement to focus on the comprehension aspect of situational awareness in order to have the system not prove detrimental under physiological factors. This further reinforces the importance of consideration of use-case, as performance metrics that may be targeted for improvement may lead to a decline in other performance metrics, and so consideration must be taken to ensure these systems do not have a meaningful detrimental impact, especially in high-risk environments such as the health or industrial sectors.

In *Chapter 6 (RQ 3)*, this thesis highlights the critical difficulty in supporting the second stage of Situational Awareness - Comprehension, and investigates new ways of defining situational awareness, based on the failure that occurs, so that support systems for these failure types can potentially be developed. From this chapter, this thesis focuses on the failure point of Semantic Comprehension, and categorization of stimuli, as an area of augmentation with the use of BCIs. To do this, this thesis presents a preliminary study based on detecting one of these failures, specifically failures in the process of Semantic Comprehension, such that future systems can potentially be developed to support users during the Comprehension stage of Situational Awareness. The results of this chapter indicate a clear biomarker located in the Frontal and Temporal-Parietal regions of the brain, which provides an avenue for classification and integration with future BCI systems.

7.2 Key Contributions

- Identification of the capability of leveraging previous experience, specifically in recreational video games, for target localization.
- Identifying the capability of increasing Situational Awareness in a state of physiological Stress using Heads-Up Displays.
- Determining the effects of Circadian Fatigue on SA, and mitigating SA loss through HUD support.
- Identifying that physiological factors can improve reaction time, however both investigated factors led to a reduction in high-level cognitive capability.
- Identifying the brain dynamics for semantic comprehension, highlighting the possibility for error detection and correction with BCI support systems.

7.3 Summary & Future Work

This thesis has outlined the current state of Situational Awareness support systems, defined research questions to further expand on the area of wearable support aides, and conducted multiple studies, including analysis and discussion of each study. This thesis found that previous experience with HUDs can be leveraged for better user experience, and HUDs based off said designs can show tangible benefits in states of stress and fatigue, however these benefits are not universal. To this end, this thesis ends with a preliminary study into the ability to detect the process of comprehension within the brain, to support integration of BCI technology for further support capabilities.

The author believes that future research into the area of BCI integration can provide significant benefits for improving SA, ranging from missed stimuli to misinterpreted information. To this end, the author believes that the experiments conducted in the virtual environment prove as good examples of how to emulate real-world situations when advanced BCI technology is currently limited to laboratory recordings.

Furthermore, the author would like to propose two major areas of future research to ensure the full capability of a real-time BCI support system. As core issues of BCI systems, the time taken to classify, as well as the robustness of the data collected, these two issues can prevent a system from working in-situ, when said system performs acceptably in a laboratory environment. To this end, the two following studies can support further development of this system.

7.3.1 Real-time Classification of Comprehension

The first core issue of a usable system, and thus the first future work to be investigated is that of real-time classification. Whilst it is possible to extract meaningful information from EEG data when processing the data with a full processing pipeline on all available channels, including independent component analysis, this is often not viable in a real-time system. Therefore, further research can be conducted on the minimum number of channels needed to achieve a meaningful classification accuracy, as well as whether components are required for a usable system. The findings from this further work would allow a system and pipeline to be developed that allows a meaningful classification accuracy to be achieved in real-time, which is the first major milestone in achieving a fully functioning BCI-based support system.

7.3.2 Dry-Sensor Based Classification

The second core issue for a BCI-based support system is that of portability and data quality. Whilst the preliminary study conducted as part of this thesis made use of a 128 channel wet-sensor based EEG system, this system is not practical for applications outside of a laboratory environment. Unfortunately, this is a common trade-off with EEG systems, where the higher fidelity data the system can produce, the less portable the system, as well as the more time required for setup. In order to have a system that is man-portable, it must be determined whether EEG signals collected by a man-portable system such as a dry-sensor based EEG system is of a high enough fidelity to be usable within classification. If current dry-sensor technology is not able to collect data with which meaningful classification accuracy can be achieved, then unfortunately BCI-based support systems may be relegated to static stations until a higher-fidelity portable system is developed. This difficulty in employing EEG-based systems has ,in fact, been noted by Rekrut et al. [224], who have currently begun undertaking research to identify if it is possible to reduce the number of EEG channels required for data collection whilst maintaining an acceptable amount of semantic data for classification, further highlighting the importance of this issue for the adoption of BCI technology.

7.3.3 Closed-Loop BCI for Situational Awareness Enhancement

With the two prior listed future works achieved, a functioning man-portable real-time system can be developed. This would allow for a closed-loop system to be developed, and further study to be conducted on the effectiveness of BCI-based SA support systems. This would be the core final future work of this thesis, as it would allow the demonstration as to the full use of a BCI-based Heads-up-display for SA support, in order to improve operator performance, as well as improve workplace safety through increased situational awareness.

APPENDIX A

APPENDIX

A.1 Appendix 1 - Equipment Used

A.1.1 Hardware

Hardware Item	Chapter(s) Used In	Reason for Use
HTC Vive Pro Eye	Chapters 4 & 5	Simulation of Realistic Environment
Heightened Plank and Safety Harness	Chapter 4	Increasing Stress Level during Experiment
Fitbit Sense Smart Watch	Chapter 5	Monitoring Participant Fatigue Level
LiveAmp 64 EEG System	Chapter 6	Amplification of EEG Signal for Data Recording
Neuroscan Quik-cap 128 EEG System	Chapter 6	Collection of EEG Data

Table A.1: Table of Hardware Equipment used in this Thesis

A.1.2 Software

Software & Version	Chapter(s) Used In	Reason for Use
Unity 2019.3f41	Chapters 3,4,5,6	Experiment Development
SPSS 25	Chapter 5	Statistical Analysis
MATLAB v2021b, v2023a	Chapters 3,4,5,6	Data and Statistical Analysis
R Studio	Chapter 5	Analysis of Longitudinal Data
REDCap Online Database	Chapter 5	Storage and Synchronisation of Longitudinal Data
Curry 7	Chapter 6	EEG Data Collection
EEGLAB v2023.1	Chapter 6	EEG Data Analysis

Table A.2: Table of Software used in this Thesis

A.2 Appendix 2 - Questionnaires Used

A.2.1 Self Assessment Manikin

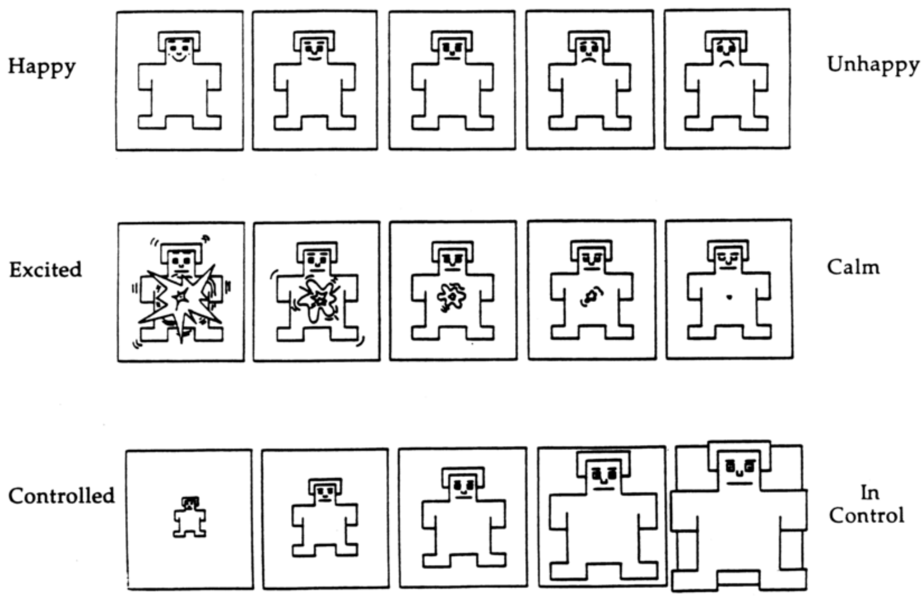
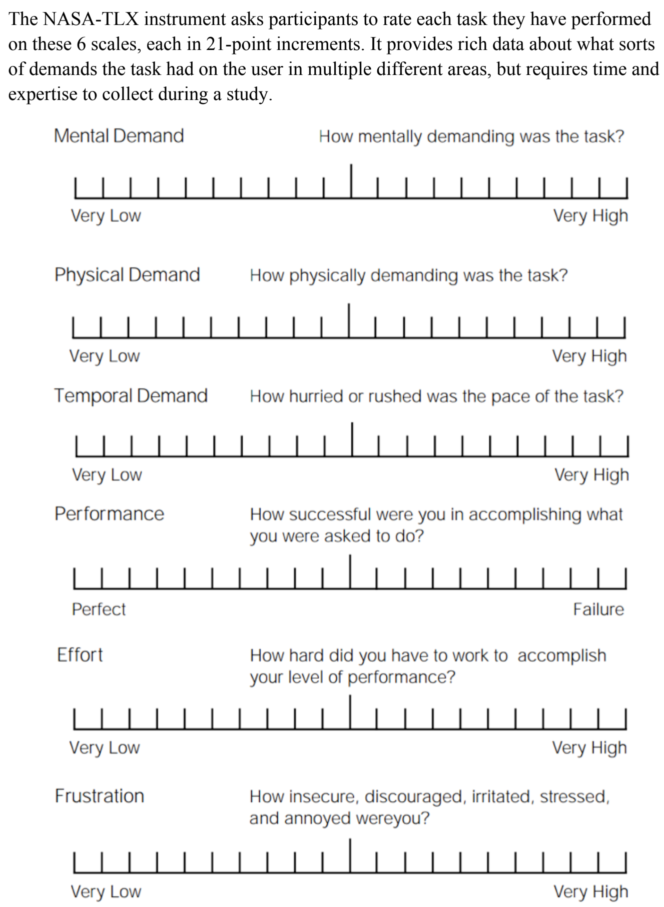


Figure A.1: The Self Assessment Manikin presented to participants

A.2.2 NASA-TLX



The NASA-TLX instrument asks participants to rate each task they have performed on these 6 scales, each in 21-point increments. It provides rich data about what sorts of demands the task had on the user in multiple different areas, but requires time and expertise to collect during a study.

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure A.2: The 21-point NASA-TLX that participants completed

A.2.3 SAGAT Questions

Plank:

- Stage 1:
  - How many buildings can you see that are taller than the one you are currently stood on? (3)
  - How many water towers can you see on the roofs of buildings? (6, 3 easily)
  - What colour is the plank? (White)
  - How many windows are in a block on the building behind you? (10)
- Stage 2:
  - Would a helicopter be able to land on the building the plank is based off (No, roof clutter would prevent it)
  - What compass direction is the helicopter facing? (230deg – SW 225 WSW 247.5)
- Stage 3:
  - Based on the helicopter’s current position and facing, which landing pad do you think it will land on? (North-West)
  - If the helicopter descended vertically, would it hit a building before landing on the ground? (Yes, barely)

Ground:

- Stage 1:
  - Which direction is the ship on the river? (South/Behind, facing towards the participant)
  - Are there any other people in the scene, and if so, where? (Yes, at the right hand side of the bridge)
  - Describe the background noise (Cars, brake noises, people noises)
  - What advertisements are on the buildings around you? (Sparkle Soda, Burgers, Vacation, Cheese, Place Your Ad Here)
- Stage 2:
  - What are the dangers of your current position? (Stood in the middle of a road, could fall off the bridge)
  - What methods of public transport are available in the city that you are aware of (Boats, Bus Stop)
- Stage 3:
  - What will happen if the ship continues down the river? (Crash into bridge as too large)
  - If a bus were to arrive at the bus stop, which direction would it be facing? (Away from the participant)

Figure A.3: The SAGAT questions presented to participants, including correct answers in brackets

Bibliography

[1] Doris Aschenbrenner et al.
"An exploration study for augmented and virtual reality enhancing situation awareness for plant teleanalysis".
In: *Proceedings of the ASME Design Engineering Technical Conference*.
Vol. 1.
American Society of Mechanical Engineers (ASME), 2017.

[2] Mica R. Endsley.
"Design and Evaluation for Situation Awareness Enhancement".
In: *Proceedings of the Human Factors Society Annual Meeting* 32.2 (Oct. 1988), pp. 97–101.

[3] Oliver W. Klaproth et al.
"Tracing Pilots' Situation Assessment by Neuroadaptive Cognitive Modeling".
In: *Frontiers in Neuroscience* 14 (Aug. 2020).

[4] Mica R. Endsley.
Toward a theory of situation awareness in dynamic systems.
Mar. 1995.

[5] Gregory Bedny and David Meister.
"Theory of Activity and Situation Awareness".
In: *International Journal of Cognitive Ergonomics* 3.1 (Jan. 1999), pp. 63–72.

[6] K. Smith and P. A. Hancock.
Situation awareness is adaptive, externally directed consciousness.
Mar. 1995.

[7] Mica R. Endsley and Debra G Jones.
Designing for situation awareness: An approach to user-centered design.
CRC Press, 2012.

[8] Toshiyuki Inagaki.

"Situation-adaptive autonomy for time-critical takeoff decisions".
In: *International Journal of Modelling and Simulation* 20.2 (2000), pp. 175–180.

[9] David C Foyle, Anthony D Andre, and Becky L Hooey.
"Situation Awareness in an Augmented Reality Cockpit : Design , Viewpoints and Cognitive Glue Problem : Airport Surface Operations".
In: *Human Computer Interaction* (2005).

[10] Richard W Pew.
"The State of Situation Awareness Measurement: Heading Toward the Next Century".
In: *Situation Awareness Analysis and Measurement*.
Ed. by Mica R. Endsley and Daniel J. Garland.
Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2000,
Pp. 33–47.

[11] Mica R. Endsley.
"SITUATION AWARENESS GLOBAL ASSESSMENT TECHNIQUE (SAGAT)."
In: *IEEE Proceedings of the National Aerospace and Electronics Conference*.
IEEE, 1988,
Pp. 789–795.

[12] Ruilin Li, Lipo Wang, and Olga Sourina.
"Subject matching for cross-subject EEG-based recognition of driver states related to situation awareness".
In: *Methods* (Apr. 2021).

[13] Lee Guan Yeo et al.
"Mobile EEG-based situation awareness recognition for air traffic controllers".
In: *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*.
Vol. 2017-Janua.
Institute of Electrical and Electronics Engineers Inc., Nov. 2017,
Pp. 3030–3035.

[14] David M. Regal, William H. Rogers, and George P. Boucek.
"Situational awareness in the commercial flight deck: Definition, measurement, and enhancement".
In: *SAE Technical Papers*.
SAE International, Oct. 1988.

[15] Delmar M Fadden, Rolf Braune, and John Wiedemann.
"Spatial displays as a means to increase pilot situational awareness".
In: *Spatial Displays and Spatial Instruments* (1993), pp. 31–35.

[16] Doris Aschenbrenner et al.
"ARTab - using Virtual and Augmented Reality Methods for an improved Situation Awareness for Telemaintenance".
In: *IFAC-PapersOnLine* 49.30 (Jan. 2016), pp. 204–209.

[17] Evelyn Rose Saus et al.
"The effect of brief situational awareness training in a police shooting simulator: An experimental study".
In: *Military Psychology* 18.SUPPL. (2006).

[18] K. S. O'Brien and D. O'Hare.
"Situational awareness ability and cognitive skills training in a complex real-world task".
In: *Ergonomics* 50.7 (July 2007), pp. 1064–1091.

[19] Francis T. Durso et al.
"En route operational errors and situation awareness".
In: *International Journal of Aviation Psychology* 8.2 (1998), pp. 177–194.

[20] Anne Sneddon, Kathryn Mearns, and Rhona Flin.
"Stress, fatigue, situation awareness and safety in offshore drilling crews".
In: *Safety Science* 56 (July 2013), pp. 80–88.

[21] Pamela S. Tsang and Michael A. Vidulich.
"Mental Workload and Situation Awareness".
In: *Handbook of Human Factors and Ergonomics*.
Hoboken, NJ, USA: John Wiley & Sons, Inc., Feb. 2006,
Pp. 243–268.

[22] Steven J. Kass, Kerstan S. Cole, and Claudia J. Stanny.
"Effects of distraction and experience on situation awareness and simulated driving".
In: *Transportation Research Part F: Traffic Psychology and Behaviour* 10.4 (July 2007), pp. 321–329.

[23] Paul M. Salmon et al.

"What really is going on? Review of situation awareness models for individuals and teams".
In: *Theoretical Issues in Ergonomics Science* 9.4 (2008), pp. 297–323.

[24] N. A. Stanton et al.
State-of-science: situation awareness in individuals, teams and systems.
Apr. 2017.

[25] Gary Klein.
"Whose Fallacies?"
In: *Journal of Cognitive Engineering and Decision Making* 9.1 (Mar. 2015), pp. 55–58.

[26] Mica R. Endsley.
"Situation awareness misconceptions and misunderstandings".
In: *Journal of Cognitive Engineering and Decision Making* 9.1 (Mar. 2015), pp. 4–32.

[27] A. Ian Glendon and Sharon Clarke.
Human Safety and Risk Management.
2018.

[28] Ken McAnally et al.
"Inference in the Wild: A Framework for Human Situation Assessment and a Case Study of Air Combat".
In: *Cognitive Science* 42.7 (Sept. 2018), pp. 2181–2204.

[29] Dan Chiappe, Thomas Z. Strybel, and Kim Phuong L. Vu.
"A situated approach to the understanding of dynamic situations".
In: *Journal of Cognitive Engineering and Decision Making* 9.1 (Mar. 2015), pp. 33–43.

[30] William Jones and Mica R. Endsley.
"A Model of Inter- and Intrateam Situational Awareness: Implications for Design, Training, and Measurement".
In: *New Trends in Cooperative Activities: Understanding System Dynamics in Complex Environments* January 2001 (2001), p. 46 67.

[31] Ulric Neisser.
Cognition and reality : principles and implications of cognitive psychology.
San Francisco: W.H. Freeman, 1976.

[32] N. A. Stanton et al.
"Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology".
In: *Ergonomics* 49.12-13 (Oct. 2006), pp. 1288–1311.

[33] Neville A. Stanton.
Distributed situation awareness.
Jan. 2016.

[34] Debra Gordon Jones and Mica R. Endsley.
"Sources of situation awareness errors in aviation".
In: *Aviation Space and Environmental Medicine* 67.6 (1996), pp. 507–512.

[35] Francis T. Durso and Scott D. Gronlund.
"Situation awareness".
In: *Handbook of applied cognition* (1999), pp. 283–314.

[36] Arien Mack and Irvin Rock.
Inattentional blindness.
MIT press, 1998.

[37] A. D. Baddelay.
"SELECTIVE ATTENTION AND PERFORMANCE IN DANGEROUS ENVIRONMENTS".
In: *British Journal of Psychology* 63.4 (1972), pp. 537–546.

[38] Frederic Charles Bartlett.
"Fatigue following highly skilled work".
In: *Proceedings of the Royal Society of London. Series B - Biological Sciences* 131.864 (May 1943), pp. 247–257.

[39] D. E. Broadbent.
"Some Effects of Noise on Visual Performance".
In: *Quarterly Journal of Experimental Psychology* 6.1 (Mar. 1954), pp. 1–5.

[40] George A. Miller.
"The magical number seven, plus or minus two: some limits on our capacity for processing information".
In: *Psychological Review* 63.2 (Mar. 1956), pp. 81–97.

[41] Nelson Cowan.
"Visual and auditory working memory capacity".

In: *Trends in Cognitive Sciences* 2.3 (Mar. 1998), pp. 77–78.

[42] Nelson Cowan.
Working Memory Capacity.
Psychology Press, Sept. 2005,
Pp. 1–225.

[43] A. D. Baddeley.
Working memory.
New York, NY, US, 1986.

[44] Carmen Sandi and M. Teresa Pinelo-Nava.
Stress and memory: Behavioral effects and neurobiological mechanisms.
2007.

[45] Andrew Hollingworth, Carrick C. Williams, and John M. Henderson.
"To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes".
In: *Psychonomic Bulletin and Review* 8.4 (2001), pp. 761–768.

[46] Andrew Hollingworth and John M. Henderson.
"Accurate visual memory for previously attended objects in natural scenes".
In: *Journal of Experimental Psychology: Human Perception and Performance* 28.1 (2002), pp. 113–136.

[47] Steven K. Feiner.
"Augmented reality: A new way of seeing".
In: *Scientific American* 286.4 (2002), p. 48.

[48] Christopher D. Wickens et al.
Engineering psychology and human performance.
Psychology Press, 2015.

[49] Chris Furmanski, Ronald Azuma, and Mike Daily.
"Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information".
In: *Proceedings - International Symposium on Mixed and Augmented Reality, ISMAR 2002*.
Institute of Electrical and Electronics Engineers Inc., 2002,
Pp. 215–224.

[50] Yung Ching Liu and Ming Hui Wen.

"Comparison of head-up display (HUD) vs. head-down display (HDD): Driving performance of commercial vehicle operators in Taiwan".
In: *International Journal of Human Computer Studies* 61.5 (Nov. 2004), pp. 679–697.

[51] Blaine Bell and Steven Feiner.
"Augmented Reality for collaborative exploration of unfamiliar environments".
In: *NSF Lake Tahoe Workshop on Collaborative Virtual Reality and Visualization (CVRV 2003)*.
2003,
Pp. 1–7.

[52] Javier Irizarry et al.
"InfoSPOT: A mobile Augmented Reality method for accessing building information through a situation awareness approach".
In: *Automation in Construction* 33 (Aug. 2013), pp. 11–23.

[53] Willem Le Roux.
"The use of augmented reality in command and control situation awareness".
In: *Scientia Militaria - South African Journal of Military Studies* 38.1 (Aug. 2011).

[54] William Losina Brandao and Marcio Sarroglia Pinho.
"Using augmented reality to improve dismounted operators' situation awareness".
In: *Proceedings - IEEE Virtual Reality*.
IEEE Computer Society, Apr. 2017,
Pp. 297–298.

[55] Eric Gans et al.
"Augmented reality technology for day/night situational awareness for the dismounted Soldier".
In: *Display Technologies and Applications for Defense, Security, and Avionics IX; and Head- and Helmet-Mounted Displays XX*.
Ed. by Daniel D. Desjardins et al.
Vol. 9470.
SPIE, May 2015,
P. 947004.

[56] Blaine Bell, Tobias Höllerer, and Steven Feiner.
"An annotated situation-awareness aid for augmented reality".

In: *UIST (User Interface Software and Technology): Proceedings of the ACM Symposium*.
New York, New York, USA: Association for Computing Machinery (ACM), 2002,
Pp. 213–216.

[57] Steven J. Henderson and Steven Feiner.
"Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret".
In: *Science and Technology Proceedings - IEEE 2009 International Symposium on Mixed and Augmented Reality, ISMAR 2009*.
2009,
Pp. 135–144.

[58] A. Y.C. Nee et al.
"Augmented reality applications in design and manufacturing".
In: *CIRP Annals - Manufacturing Technology* 61.2 (2012), pp. 657–679.

[59] Jeff Johnson.
Designing with the Mind in Mind, Second Edition: Simple Guide to Understanding User Interface Design Guidelines.
2nd ed.
Morgan Kaufmann, 2014.

[60] Ok Choon Park and Stuart S. Gittelman.
"Dynamic characteristics of mental models and dynamic visual displays".
In: *Instructional Science* 23.5-6 (Nov. 1995), pp. 303–320.

[61] Michael Haller, Mark Billinghurst, and Bruce Thomas.
Emerging Technologies of Augmented Reality: Interfaces and Design.
1st ed.
Igi Global, 2006.

[62] David Pinelle, Nelson Wong, and Tadeusz Stach.
"Heuristic evaluation for games: Usability principles for video game design".
In: *Conference on Human Factors in Computing Systems - Proceedings*.
New York, New York, USA: ACM Press, 2008,
Pp. 1453–1462.

[63] Uwe Gruenefeld et al.
"EyeSee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality".

In: *SUI 2017 - Proceedings of the 2017 Symposium on Spatial User Interaction*. Association for Computing Machinery, Inc, Oct. 2017, Pp. 109–118.

[64] Felix Bork et al. “Towards efficient visual guidance in limited field-of-view head-mounted displays”. In: *IEEE Transactions on Visualization and Computer Graphics* 24.11 (Nov. 2018), pp. 2983–2992.

[65] Eric D. Ragan et al. “Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task”. In: *IEEE Transactions on Visualization and Computer Graphics* 21.7 (July 2015), pp. 794–807.

[66] Alexander Marquardt et al. “Comparing Non-Visual and Visual Guidance Methods for Narrow Field of View Augmented Reality Displays”. In: *IEEE Transactions on Visualization and Computer Graphics* 26.12 (Dec. 2020), pp. 3389–3401.

[67] Difeng Yu et al. “Design and Evaluation of Visualization Techniques of Off-Screen and Occluded Targets in Virtual Reality Environments”. In: *IEEE Transactions on Visualization and Computer Graphics* 26.9 (2020), pp. 2762–2774.

[68] Torben Schinke, Niels Henze, and Susanne Boll. “Visualization of off-screen objects in mobile augmented reality”. In: *ACM International Conference Proceeding Series*. 2010, Pp. 313–316.

[69] Bruce Thomas et al. “A wearable computer system with augmented reality to support terrestrial navigation”. In: *International Symposium on Wearable Computers, Digest of Papers*. Vol. 1998-Octob. IEEE Computer Society, 1998, Pp. 168–171.

[70] Ágnes Sz≈ ë$ll \approx$ ë$siEtal.^{\frown}$.
"Acute stress affects prospective memory functions via associative memory processes".
In: *Acta Psychologica* 182 (Jan. 2018), pp. 82–90.

[71] Alexandra L. Shilton, Robin Laycock, and Sheila G. Crewther.
"Different effects of trait and state anxiety on global-local visual processing following acute stress".
In: *Cognition, Brain, Behavior. An Interdisciplinary Journal* 23.3 (2019), pp. 155–170.

[72] Eduardo Salas, Aaron S Dietz, and Richard W Pew.
"The State of Situation Awareness Measurement: Heading Toward the Next Century".
In: *Situational Awareness*.
Routledge, Nov. 2018,
Pp. 459–474.

[73] Mica R. Endsley and Daniel J Garland.
Situation awareness analysis and measurement.
CRC Press, 2000.

[74] Neville A. Stanton et al.
"Use of Highways in the Sky and a virtual pad for landing Head Up Display symbology to enable improved helicopter pilots situation awareness and workload in degraded visual conditions".
In: *Ergonomics* 62.2 (Feb. 2019), pp. 255–267.

[75] Neville A. Stanton et al.
"Seeing through the mist: an evaluation of an iteratively designed head-up display, using a simulated degraded visual environment, to facilitate rotary-wing pilot situation awareness and workload".
In: *Cognition, Technology and Work* 22.3 (Aug. 2020), pp. 549–563.

[76] Edith Fischer, Richard F Haines, and Toni A Price.
"COGNITIVE ISSUES IN HEAD-UP DISPLAYS."
In: *NASA Technical Paper* 1711 (1980).

[77] Juhee Park and Woojin Park.
Functional requirements of automotive head-up displays: A systematic review of literature from 1994 to present.

Apr. 2019.

[78] Saroj K.L. Lal and Ashley Craig.
"Driver fatigue: Electroencephalography and psychological assessment".
In: *Psychophysiology* 39.3 (2002), pp. 313–321.

[79] Kuan Chih Huang et al.
"The effects of different fatigue levels on brain‚Äìbehavior relationships in driving".
In: *Brain and Behavior* 9.12 (2019).

[80] Tien Thong Nguyen Do, Yu Kai Wang, and Chin Teng Lin.
"Increase in Brain Effective Connectivity in Multitasking but not in a High-Fatigue State".
In: *IEEE Transactions on Cognitive and Developmental Systems* 13.3 (2021), pp. 566–574.

[81] Hans Henrik Sievertsen, Francesca Gino, and Marco Piovesan.
"Cognitive fatigue influences students' performance on standardized tests".
In: *Proceedings of the National Academy of Sciences of the United States of America* 113.10 (Mar. 2016), pp. 2621–2624.

[82] André Fonseca et al.
"Brain network changes in fatigued drivers: A longitudinal study in a real-world environment based on the effective connectivity analysis and actigraphy data".
In: *Frontiers in Human Neuroscience* 12 (Nov. 2018), p. 418.

[83] Frédéric Dehais et al.
"Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid fNIRS-EEG Passive BCI".
In: *Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018*.
Institute of Electrical and Electronics Engineers Inc., Jan. 2019,
Pp. 544–549.

[84] Hong Wang et al.
"Real-time EEG-based detection of fatigue driving danger for accident prediction".
In: *International Journal of Neural Systems* 25.2 (Mar. 2015).

[85] Yu Te Wang et al.

"Developing an EEG-based on-line closed-loop lapse detection and mitigation system".
In: *Frontiers in Neuroscience* 8.OCT (Jan. 2014), p. 321.

[86] Anwesha Sengupta et al.
"EEG synchronization and brain networks: A case study in fatigue".
In: *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems, MedCom 2014*.
Institute of Electrical and Electronics Engineers Inc., 2014,
Pp. 278–282.

[87] Raphaelle N. Roy et al.
"Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI".
In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*.
IEEE, July 2013,
Pp. 6607–6610.

[88] Guillermo Borragán et al.
"Decreased prefrontal connectivity parallels cognitive fatigue-related performance decline after sleep deprivation. An optical imaging study".
In: *Biological Psychology* 144 (May 2019), pp. 115–124.

[89] Scott Makeig, Tzyy Ping Jung, and Terrence J. Sejnowski.
"Awareness during drowsiness: Dynamics and electrophysiological correlates".
In: *Canadian Journal of Experimental Psychology* 54.4 (2000), pp. 266–273.

[90] Anne Sneddon, Kathryn Mearns, and Rhona Flin.
"Situation awareness and safety in offshore drill crews".
In: *Cognition, Technology and Work* 8.4 (Nov. 2006), pp. 255–267.

[91] Michael D. Matthews et al.
"A comparison of expert ratings and self-assessments of situation awareness during a combat fatigue course".
In: *Military Psychology* 23.2 (Mar. 2011), pp. 125–136.

[92] National Heavy Vehicle Regulator.
"Basic Fatigue Management Accreditation Guide".
In: February (2014).

[93] Micah K. Wilson, Luke Strickland, and Timothy Ballard.
"FIPS: An R Package for Biomathematical Modelling of Human Fatigue Related Impairment".
In: *Journal of Open Source Software* 5.51 (July 2020), p. 2340.

[94] D.J. Weintraub and M Ensing.
Human Factors Issues in Head-Up Display Design: The Book of HUD.
Tech. rep.
1992,
Pp. 1–213.

[95] Dale Richards and Phillip Lamb.
"Functional Symbology - Evaluation of task-specific Head-Up Display information for use on a commercial flight deck".
In:
2016.

[96] Steven Henderson and Steven Feiner.
"Exploring the benefits of augmented reality documentation for maintenance and repair".
In: *IEEE Transactions on Visualization and Computer Graphics* 17.10 (2011), pp. 1355–1368.

[97] Ali Farhadi et al.
"Describing objects by their attributes".
In: *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*.
Vol. 2009 IEEE.
2009,
Pp. 1778–1785.

[98] Peter Garrard et al.
"Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts".
In: *Cognitive Neuropsychology* 18.2 (2001), pp. 125–174.

[99] Alexander M. Chan et al.
"Decoding word and category-specific spatiotemporal representations from MEG and EEG".
In: *NeuroImage* 54.4 (Feb. 2011), pp. 3028–3039.

[100] Jakub M. Szewczyk and Herbert Schriefers.
"Is animacy special?: ERP correlates of semantic violations and animacy violations in sentence processing".
In: *Brain Research* 1368 (Jan. 2011), pp. 208–221.

[101] Jill Weckerly and Marta Kutas.
"An electrophysiological analysis of animacy effects in the processing of object relative sentences".
In: *Psychophysiology* 36.5 (Sept. 1999), pp. 559–570.

[102] Lin Wang et al.
"Neural evidence for the prediction of animacy features during language comprehension: Evidence from MEG and EEG representational similarity analysis".
In: *Journal of Neuroscience* 40.16 (Apr. 2020), pp. 3278–3291.

[103] Thomas Carlson et al.
"Representational dynamics of object vision: The first 1000 ms".
In: *Journal of Vision* 13.10 (Aug. 2013), pp. 1–1.

[104] Marzia De Lucia et al.
"Auditory perceptual decision-making based on semantic categorization of environmental sounds".
In: *NeuroImage* 60.3 (Apr. 2012), pp. 1704–1715.

[105] João M. Correia et al.
"EEG decoding of spoken words in bilingual listeners: From words to language invariant semantic-conceptual representations".
In: *Frontiers in Psychology* 6.FEB (Feb. 2015), p. 71.

[106] Juan Linde-Domingo et al.
"Evidence that neural information flow is reversed between object perception and object reconstruction from memory".
In: *Nature Communications* 10.1 (Dec. 2019), p. 179.

[107] Wilma A. Bainbridge, Elizabeth H. Hall, and Chris I. Baker.
"Distinct Representational Structure and Localization for Visual Encoding and Recall during Visual Imagery".
In: *Cerebral Cortex* 31.4 (Apr. 2021), pp. 1898–1913.

[108] Michelle E. Costanzo et al.

"Spatial and temporal features of superordinate semantic processing studied with fMRI and EEG".
In: *Frontiers in Human Neuroscience* 7.JUN (June 2013), p. 293.

[109] Alexander G. Huth et al.
"Natural speech reveals the semantic maps that tile human cerebral cortex".
In: *Nature* 532.7600 (Apr. 2016), pp. 453–458.

[110] Brian Murphy et al.
"EEG decoding of semantic category reveals distributed representations for single concepts".
In: *Brain and Language* 117.1 (Apr. 2011), pp. 12–22.

[111] Richard D. Sperber et al.
"Semantic priming effects on picture and word processing".
In: *Memory & Cognition* 7.5 (Sept. 1979), pp. 339–345.

[112] Jeffrey R. Binder et al.
"Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies".
In: *Cerebral Cortex* 19.12 (Dec. 2009), pp. 2767–2796.

[113] W. Klimesch, H. Schimke, and J. Schwaiger.
"Episodic and semantic memory: an analysis in the EEG theta and alpha band".
In: *Electroencephalography and Clinical Neurophysiology* 91.6 (Dec. 1994), pp. 428–441.

[114] W. Klimesch et al.
"Event-related desynchronization in the alpha band and the processing of semantic information".
In: *Cognitive Brain Research* 6.2 (Oct. 1997), pp. 83–94.

[115] Mandy J. Maguire, Matthew R. Brier, and Thomas C. Ferree.
"EEG theta and alpha responses reveal qualitative differences in processing taxonomic versus thematic semantic relationships".
In: *Brain and Language* 114.1 (July 2010), pp. 16–25.

[116] Peter Hagoort, Colin M. Brown, and Tamara Y. Swaab.
"Lexical-semantic event-related potential effects in patients with left hemisphere lesions and aphasia, and patients with right hemisphere lesions without aphasia".
In: *Brain* 119.2 (1996), pp. 627–649.

[117] P. Arico et al.
Passive BCI beyond the lab: Current trends and future directions.
Aug. 2018.

[118] Benjamin Blankertz et al.
The Berlin brain-computer interface: Progress beyond communication and control.
Nov. 2016.

[119] Chin Teng Lin and Tien Thong Nguyen Do.
"Direct-Sense Brain-Computer Interfaces and Wearable Computers".
In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 51.1 (Jan. 2021), pp. 298–312.

[120] Nader Alharbi.
"A novel approach for noise removal and distinction of EEG recordings".
In: *Biomedical Signal Processing and Control* 39 (Jan. 2018), pp. 23–33.

[121] Chi Qin Lai et al.
"Artifacts and noise removal for electroencephalogram (EEG): A literature review".
In: *ISCAIE 2018 - 2018 IEEE Symposium on Computer Applications and Industrial Electronics*.
Institute of Electrical and Electronics Engineers Inc., July 2018,
Pp. 326–332.

[122] Samir Sangani, Anouk Lamontagne, and Joyce Fung.
"Cortical mechanisms underlying sensorimotor enhancement promoted by walking with haptic inputs in a virtual environment".
In: *Progress in Brain Research*.
Vol. 218.
Elsevier B.V., Jan. 2015,
Pp. 313–330.

[123] Siamac Fazli et al.
"Enhanced performance by a hybrid NIRS-EEG brain computer interface".
In: *NeuroImage* 59.1 (Jan. 2012), pp. 519–529.

[124] Rihui Li et al.
"Enhancing performance of a hybrid EEG-fNIRS system using channel selection and early temporal features".
In: *Frontiers in Human Neuroscience* 11 (Sept. 2017), p. 462.

[125] Sheng Ge et al.
"A Brain-Computer Interface Based on a Few-Channel EEG-fNIRS Bimodal System".
In: *IEEE Access* 5 (2017), pp. 208–218.

[126] Xuxian Yin et al.
"A hybrid BCI based on EEG and fNIRS signals improves the performance of decoding motor imagery of both force and speed of hand clenching".
In: *Journal of Neural Engineering* 12.3 (June 2015), p. 036004.

[127] Mustafa A.H. Hasan, Muhammad U. Khan, and Deepti Mishra.
"A Computationally Efficient Method for Hybrid EEG-fNIRS BCI Based on the Pearson Correlation".
In: *BioMed Research International* 2020 (2020).

[128] Mansi Sharma et al.
"Distinguishing Target and Non-Target Fixations with EEG and Eye Tracking in Realistic Visual Scenes".
In: *ACM International Conference Proceeding Series*.
Association for Computing Machinery, Nov. 2024,
Pp. 459–468.

[129] Yufeng Ke et al.
"An EEG-Based mental workload estimator trained on working memory task can work well under simulated Multi-Attribute task".
In: *Frontiers in Human Neuroscience* 8.SEP (Sept. 2014), p. 703.

[130] Samantha A. Sprague, Matthew T. McBee, and Eric W. Sellers.
"The effects of working memory on brain-computer interface performance".
In: *Clinical Neurophysiology* 127.2 (Feb. 2016), pp. 1331–1341.

[131] Joshua J. LaRocque et al.
"Decoding attended information in short-term memory: An EEG study".
In: *Journal of Cognitive Neuroscience* 25.1 (Jan. 2013), pp. 127–142.

[132] Seolhwa Lee et al.
"Comparing Programming Language Comprehension between Novice and Expert Programmers Using EEG Analysis".
In: *Proceedings - 2016 IEEE 16th International Conference on Bioinformatics and Bioengineering, BIBE 2016*.
Institute of Electrical and Electronics Engineers Inc., Dec. 2016,

Pp. 350–355.

[133] Radoslaw Martin Cichy and Dimitrios Pantazis.
"Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space".
In: *NeuroImage* 158 (Sept. 2017), pp. 441–454.

[134] Mehdi Behroozi, Mohammad Reza Daliri, and Babak Shekarchi.
"EEG phase patterns reflect the representation of semantic categories of objects".
In: *Medical and Biological Engineering and Computing* 54.1 (Jan. 2016), pp. 205–221.

[135] Fabien Perrin and Luis García-Larrea.
"Modulation of the N400 potential during auditory phonological/semantic interaction".
In: *Cognitive Brain Research* 17.1 (June 2003), pp. 36–47.

[136] Majid Khalili Ardali et al.
"Semantic and BCI-performance in completely paralyzed patients: Possibility of language attrition in completely locked in syndrome".
In: *Brain and Language* 194 (July 2019), pp. 93–97.

[137] Maurice Rekrut et al.
"Decoding Semantic Categories from EEG Activity in Silent Speech Imagination Tasks".
In: *9th IEEE International Winter Conference on Brain-Computer Interface, BCI 2021*.
Institute of Electrical and Electronics Engineers Inc., Feb. 2021.

[138] A J Wills et al.
"Predictive learning, prediction errors, and attention: Evidence from event-related potentials and eye tracking".
In: *Journal of Cognitive Neuroscience* 19.5 (2007), pp. 843–854.

[139] Jonathan E. Robinson et al.
"Dose-dependent modulation of the visually evoked N1/N170 by perceptual surprise: a clear demonstration of prediction-error signalling".
In: *European Journal of Neuroscience* 52.11 (2020), pp. 4442–4452.

[140] James F. Cavanagh et al.

"Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation".
In: *Cerebral Cortex* 22.11 (Nov. 2012), pp. 2575–2586.

[141] Rolf Verleger et al.
"On why targets evoke P3 components in prediction tasks: Drawing an analogy between prediction and matching tasks".
In: *Frontiers in Human Neuroscience* 11 (Oct. 2017), p. 497.

[142] Pouya Ahmadian, Stefano Cagnoni, and Luca Ascari.
"How capable is non-invasive EEG data of predicting the next movement? a mini review".
In: *Frontiers in Human Neuroscience* 7.MAR (Mar. 2013), p. 124.

[143] Yajing Si et al.
"Predicting individual decision-making responses based on single-trial EEG".
In: *NeuroImage* 206 (Feb. 2020), p. 116333.

[144] F. Lotte et al.
A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update.
2018.

[145] Fabien Lotte and Raphaëlle N. Roy.
"Brain-computer interface contributions to neuroergonomics".
In: *Neuroergonomics: The Brain at Work and in Everyday Life*.
Elsevier, Jan. 2018,
Pp. 43–48.

[146] F. Lotte et al.
A review of classification algorithms for EEG-based brain-computer interfaces.
June 2007.

[147] Irina Simanova et al.
"Identifying object categories from event-related EEG: Toward decoding of conceptual representations".
In: *PLoS ONE* 5.12 (Dec. 2010). Ed. by Hans P. Op. de Beeck, e14465.

[148] Maurice Rekrut et al.
"Decoding Semantic Categories from EEG Activity in Object-Based Decision Tasks".

In: *8th International Winter Conference on Brain-Computer Interface, BCI 2020*. Institute of Electrical and Electronics Engineers Inc., Feb. 2020.

[149] Yichen Yu et al.
"A Novel Linguistic Brain-Computer Interface Based on an Improved EEGNet".
In: *Lecture Notes in Electrical Engineering*.
Vol. 1396 LNEE.
Springer, Singapore, 2025,
Pp. 79–89.

[150] Zhaoheng Ni et al.
"Confused or not confused?: Disentangling Brain activity from EEG data using Bidirectional LSTM Recurrent Neural Networks".
In: *ACM-BCB 2017 - Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*.
New York, NY, USA: Association for Computing Machinery, Inc, Aug. 2017,
Pp. 241–246.

[151] R. Shively, Connie Brasil, and Susan Flaherty.
"Alternative UAV Sensor Control: Leveraging Gaming Skill".
In: *2007 International Symposium on Aviation Psychology* (Jan. 2007).

[152] Brian T Schreiber, Don R Lyon, and Elizabeth L Martin.
Impact of prior flight experience on learning Predator UAV operator skills.
Tech. rep. February.
2002.

[153] Amy L Alexander et al.
"From Gaming to Training : A Review of Studies on Fidelity , Immersion , Presence , and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games".
In: *DARWARS Training Impact Group* 5.November (2005), pp. 1–14.

[154] Erik Fagerholt and Magnus Lorentzon.
"Beyond the HUD. User Interfaces for Increased Player Immersion in FPS Games".
In: *Chalmers University* (2009), p. 124.

[155] Piotr Majdak, Matthew J. Goupell, and Bernhard Laback.
"3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training".
In: *Attention, Perception, and Psychophysics* 72.2 (Feb. 2010), pp. 454–469.

[156] Matthias Frank et al.
"Flexible and intuitive pointing method for 3-D auditory localization experiments".
In: *Proceedings of the AES International Conference*.
Audio Engineering Society, June 2011,
Pp. 145–153.

[157] B. Bowman, N. Elmqvist, and T. J. Jankun-Kelly.
"Toward Visualization for Games: Theory, Design Space, and Patterns".
In: *IEEE Transactions on Visualization and Computer Graphics* 18.11 (Nov. 2012), pp. 1956–1968.

[158] Veronica Zammitto.
"Visualization techniques in video games".
In: *Electronic Visualisation and the Arts (EVA 2008)* (2008), pp. 267–276.

[159] Yanqiu Tian et al.
"A Comparison of Common Video Game versus Real-World Heads-Up-Display Designs for the Purpose of Target Localization and Identification".
In: *Proceedings - 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct, ISMAR-Adjunct 2021*.
Institute of Electrical and Electronics Engineers Inc., 2021,
Pp. 228–233.

[160] C Shawn Green and Daphne Bavelier.
"Action video game modifies visual selective attention".
In: *Nature* 423.6939 (2003), pp. 534–537.

[161] Magdalena Kowal et al.
"Different cognitive abilities displayed by action video gamers and non-gamers".
In: *Computers in Human Behavior* 88 (2018), pp. 255–262.

[162] Mona Moisala et al.
"Gaming is related to enhanced working memory performance and task-related cortical activity".
In: *Brain Research* 1655 (2017), pp. 204–215.

[163] Otto Waris et al.
"Video gaming and working memory: A large-scale cross-sectional correlative study".
In: *Computers in human behavior* 97 (2019), pp. 94–103.

[164] Evan D Graham and Christine L MacKenzie.
"Physical versus virtual pointing".
In: *Proceedings of the SIGCHI conference on Human factors in computing systems*.
1996,
Pp. 292–299.

[165] Tovi Grossman and Ravin Balakrishnan.
"Pointing at trivariate targets in 3D environments".
In: *Proceedings of the SIGCHI conference on Human factors in computing systems*.
2004,
Pp. 447–454.

[166] Christopher D Wickens and Jeffry Long.
"Object versus space-based models of visual attention: Implications for the design of head-up displays."
In: *Journal of Experimental Psychology: Applied* 1.3 (1995), p. 179.

[167] Eric Gans et al.
"Augmented reality technology for day/night situational awareness for the dismounted Soldier".
In: *Display technologies and applications for defense, security, and avionics IX; and head-and helmet-mounted displays XX*.
Vol. 9470.
International Society for Optics and Photonics. 2015,
P. 947004.

[168] Loïc Caroux and Katherine Isbister.
"Influence of head-up displays,Äô characteristics on user experience in video games".
In: *International Journal of Human-Computer Studies* 87 (2016), pp. 65–79.

[169] Margaret M. Bradley and Peter J. Lang.
"Measuring emotion: The self-assessment manikin and the semantic differential".
In: *Journal of Behavior Therapy and Experimental Psychiatry* 25.1 (Mar. 1994), pp. 49–59.

[170] Matthew Lombard et al.
"Presence and television the role of screen size".
In: *Human Communication Research* 26.1 (2000), pp. 75–98.

[171] Franz Faul et al.

"Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses".
In: *Behavior Research Methods* 41.4 (2009), pp. 1149–1160.

[172] Kelly A. Burke, David J. Wing, and Timothy Lewis.
"Pilot subjective assessments during an investigation of separation function allocation using a human-in-the-loop simulation".
In: *2013 Aviation Technology, Integration, and Operations Conference*.
2013.

[173] Jessica Sänger et al.
"The influence of acute stress on attention mechanisms and its electrophysiological correlates".
In: *Frontiers in Behavioral Neuroscience* 8.OCT (Oct. 2014), p. 353.

[174] Lisa Wirz and Lars Schwabe.
"Prioritized attentional processing: Acute stress, memory and stimulus emotionality facilitate attentional disengagement".
In: *Neuropsychologia* 138 (Feb. 2020), p. 107334.

[175] Grant S. Shields et al.
"Mild acute stress improves response speed without impairing accuracy or interference control in two selective attention tasks: Implications for theories of stress and cognition".
In: *Psychoneuroendocrinology* 108 (Oct. 2019), pp. 78–86.

[176] Julia Woodward et al.
"Examining the Presentation of Information in Augmented Reality Headsets for Situational Awareness".
In: *ACM International Conference Proceeding Series*.
New York, NY, USA: Association for Computing Machinery, Sept. 2020,
Pp. 1–5.

[177] Jason Wuertz et al.
"A design framework for awareness cues in distributed multiplayer games".
In: *Conference on Human Factors in Computing Systems - Proceedings*.
Vol. 2018-April.
New York, NY, USA: Association for Computing Machinery, Apr. 2018,
Pp. 1–14.

[178] Jordan Allspaw, Lilia Heinold, and Holly A. Yanco.

"Design of Virtual Reality for Humanoid Robots with Inspiration from Video Games".
In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
Vol. 11575 LNCS.
Springer Verlag, July 2019,
Pp. 3–18.

[179] Talia Lavie and Joachim Meyer.
"Benefits and costs of adaptive user interfaces".
In: *International Journal of Human Computer Studies* 68.8 (Aug. 2010), pp. 508–524.

[180] Dragos F. Sburlana et al.
"Adaptive interactive displaying system for in-vehicle use".
In: *Procedia Computer Science*.
Vol. 176.
Elsevier, Jan. 2020,
Pp. 195–204.

[181] Henrik Detjen et al.
"Investigating the Influence of Gaze- and Context-Adaptive Head-up Displays on Take-Over Requests".
In: *Main Proceedings - 14th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2022*.
Association for Computing Machinery, Inc, Sept. 2022,
Pp. 108–118.

[182] Huicong Fang et al.
"The role of phonological loop and visuospatial sketchpad in virtual maze wayfinding".
In: *Journal of Environmental Psychology* 67 (Feb. 2020), p. 101378.

[183] Alan Baddeley.
"Working Memory: An Overview".
In: *Working Memory and Education*.
Academic Press, Jan. 2006,
Pp. 1–31.

[184] Zahra Jafari, Bryan E. Kolb, and Majid H. Mohajerani.

Effect of acute stress on auditory processing: A systematic review of human studies. Jan. 2017.

[185] Nils Kohn, Erno J. Hermans, and Guillén Fernández.
"Cognitive benefit and cost of acute stress is differentially modulated by individual brain state".
In: *Social Cognitive and Affective Neuroscience* 12.7 (July 2017), pp. 1179–1187.

[186] Robert Dudley et al.
"The effect of arousal on auditory threat detection and the relationship to auditory hallucinations".
In: *Journal of Behavior Therapy and Experimental Psychiatry* 45.3 (Sept. 2014), pp. 311–318.

[187] Robert M. Yerkes and John D. Dodson.
"The relation of strength of stimulus to rapidity of habit-formation".
In: *Journal of Comparative Neurology and Psychology* 18.5 (Nov. 1908), pp. 459–482.

[188] James L. Szalma.
"Workload and stress in vigilance: the impact of display format and task type."
In: *The American journal of psychology* 124.4 (Dec. 2011), pp. 441–454.

[189] Nicholas J. Cepeda et al.
"Spacing effects in learning: A temporal ridgeline of optimal retention".
In: *Psychological Science* 19.11 (Nov. 2008), pp. 1095–1102.

[190] Torbjörn Åkerstedt and Mats Gillberg.
"Subjective and objective sleepiness in the active individual".
In: *International Journal of Neuroscience* 52.1-2 (1990), pp. 29–37.

[191] Azmeh Shahid et al.
"Karolinska Sleepiness Scale (KSS)".
In: *STOP, THAT and One Hundred Other Sleep Scales*.
Springer, New York, NY, 2011,
Pp. 209–210.

[192] Kathryn A. Lee, Gregory Hicks, and German Nino-Murcia.
"Validity and reliability of a scale to assess fatigue".
In: *Psychiatry Research* 36.3 (Mar. 1991), pp. 291–298.

[193] Daniel J. Buysse et al.

"The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research".
In: *Psychiatry Research* 28.2 (May 1989), pp. 193–213.

[194] Michael Scott Evans, Daniel Harborne, and Andrew P. Smith.
"Developing an Objective Indicator of Fatigue: An Alternative Mobile Version of the Psychomotor Vigilance Task (m-PVT)".
In: *Communications in Computer and Information Science*.
Vol. 1012.
Springer Verlag, 2019,
Pp. 49–71.

[195] Lucia Arsintescu et al.
"Validation of a touchscreen psychomotor vigilance task".
In: *Accident Analysis and Prevention* 126 (May 2019), pp. 173–176.

[196] Michael Ingre et al.
"Validating and extending the three process model of alertness in airline operations".
In: *PLoS ONE* 9.10 (Oct. 2014), e108679.

[197] HTC.
VIVE 3DSP Audio SDK Unity Plugin ,Äî VIVE 3DSP Audio SDK v 1.2.8 documentation.
2023.

[198] Sylvain Choisel and Karin Zimmer.
"A Pointing Technique with Visual Feedback for Sound Source Localization Experiments".
In: *Audio Engineering Society*.
Audio Engineering Society, Oct. 2003.

[199] Hélène Bahu et al.
"Comparison of different egocentric pointing methods for 3D sound localization experiments".
In: *Acta Acustica united with Acustica* 102.1 (Jan. 2016), pp. 107–118.

[200] Christopher D. Wickens and Amy L. Alexander.
"Attentional Tunneling and Task Management in Synthetic Vision Displays".
In: *The International Journal of Aviation Psychology* 19.2 (Mar. 2009), pp. 182–199.

[201] Brandon Victor Syiem et al.
"Impact of task on attentional tunneling in handheld augmented reality".
In: *Conference on Human Factors in Computing Systems - Proceedings*.
Association for Computing Machinery, May 2021.

[202] Mark A. Steadman et al.
"Short-term effects of sound localization training in virtual reality".
In: *Scientific Reports* 9.1 (2019).

[203] Renata M. Heilman et al.
"Emotion Regulation and Decision Making Under Risk and Uncertainty".
In: *Emotion* 10.2 (Apr. 2010), pp. 257–265.

[204] Mickaël Causse et al.
"The effects of emotion on pilot decision-making: A neuroergonomic approach to aviation safety".
In: *Transportation Research Part C: Emerging Technologies* 33 (Aug. 2013), pp. 272–281.

[205] In Soo Lee et al.
"Number of lapses during the psychomotor vigilance task as an objective measure of fatigue".
In: *Journal of Clinical Sleep Medicine* 6.2 (Apr. 2010), pp. 163–168.

[206] Micah K. Wilson et al.
"Understanding fatigue in a naval submarine: Applying biomathematical models and workload measurement in an intensive longitudinal design".
In: *Applied Ergonomics* 94 (July 2021), p. 103412.

[207] Michael David Wilson et al.
"The next generation of fatigue prediction models: evaluating current trends in biomathematical modelling".
In: *Theoretical Issues in Ergonomics Science* (2022).

[208] William R. McMahon et al.
"The wake maintenance zone shows task dependent changes in cognitive function following one night without sleep".
In: *Sleep* 41.10 (Oct. 2018).

[209] David Grogna et al.

"The impact of drowsiness on in-vehicle human-machine interaction with head-up and head-down displays".
In: *Multimedia Tools and Applications* 77.21 (Nov. 2018), pp. 27807–27827.

[210] Yung Ching Liu and Tsun Ju Wu.
"Fatigued driver's driving behavior and cognitive task performance: Effects of road environments and road environment changes".
In: *Safety Science* 47.8 (Oct. 2009), pp. 1083–1089.

[211] Julia A. Shekleton et al.
"Improved neurobehavioral performance during the wake maintenance zone".
In: *Journal of Clinical Sleep Medicine* 9.4 (2013), pp. 353–362.

[212] Jan de Zeeuw et al.
"The alerting effect of the wake maintenance zone during 40 hours of sleep deprivation".
In: *Scientific Reports* 8.1 (July 2018), pp. 1–11.

[213] Terry McMorris et al.
Cognitive fatigue effects on physical performance: A systematic review and meta-analysis.
May 2018.

[214] Eduardo Rosa et al.
"Effects of Fatigue on Cognitive Performance in Long-Duration Simulated Flight Missions".
In: *Aviation Psychology and Applied Human Factors* 10.2 (Oct. 2020), pp. 82–93.

[215] Roger M. Enoka and Jacques Duchateau.
"Translating fatigue to human performance".
In: *Medicine and Science in Sports and Exercise* 48.11 (Nov. 2016), pp. 2228–2238.

[216] Martin Behrens et al.
Fatigue and Human Performance: An Updated Framework.
Oct. 2023.

[217] Trudy Mallinson et al.
"Giving meaning to measure: Linking self-reported fatigue and function to performance of everyday activities".
In: *Journal of Pain and Symptom Management* 31.3 (Mar. 2006), pp. 229–241.

[218] Guillermo Borragan et al.

"Cognitive fatigue facilitates procedural sequence learning".
In: *Frontiers in Human Neuroscience* 10.MAR2016 (Mar. 2016), p. 179603.

[219] Chao Wang et al.
"Compensatory neural activity in response to cognitive fatigue".
In: *Journal of Neuroscience* 36.14 (Apr. 2016), pp. 3919–3924.

[220] Philippa Gander, Curt Graeber, and Gregory Belenky.
"Operator fatigue: Implications for human-machine interaction".
In: *The Handbook of Human-Machine Interaction: A Human-Centered Design Approach* (Nov. 2011), pp. 365–382.

[221] D Browne.
Adaptive User Interfaces.
Computers and People Series.
Elsevier Science, 2016.

[222] Jamil Hussain et al.
"Model-based adaptive user interface based on context and user experience evaluation".
In: *Journal on Multimodal User Interfaces* 12.1 (Mar. 2018), pp. 1–16.

[223] Christos N. Moridis et al.
"Using EEG Frontal Asymmetry to Predict IT User's Perceptions Regarding Usefulness, Ease of Use and Playfulness".
In: *Applied Psychophysiology Biofeedback* 43.1 (Mar. 2018), pp. 1–11.

[224] Maurice Rekrut et al.
"How low can you go: evaluating electrode reduction methods for EEG-based speech imagery BCIs".
In: *Frontiers in Neuroergonomics* 6 (July 2025), p. 1578586.