2024

Reduction of cognitive load in immersive virtual reality with multisensory cues

Daniel Wilson

Recommended Citation
Wilson, Daniel, "Reduction of cognitive load in immersive virtual reality with multisensory cues" (2024). *Theses*. 673.
https://louis.uah.edu/uah-theses/673

REDUCTION OF COGNITIVE LOAD IN IMMERSIVE VIRTUAL REALITY WITH MULTISENSORY CUES

Daniel Wilson

A THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts
in
The Department of Psychology
to
The Graduate School
of
The University of Alabama in Huntsville
May 2024

Approved by:

Dr. Lisa Vangsness, Research Advisor & Committee Chair
Dr. Kristen Weger, Committee Member
Dr. Nathan Tenhundfeld, Committee Member
Dr. Nathan Tenhundfeld, Department Chair
Dr. Sean Lane, College Dean
Dr. Jon Hakkila, Graduate Dean

Abstract

REDUCTION OF COGNITIVE LOAD IN IMMERSIVE VIRTUAL REALITY WITH MULTISENSORY CUES

Daniel Wilson

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Arts

Psychology

The University of Alabama in Huntsville
May 2024

This study investigated the effects of multimodal interfaces and immersive virtual environments on cognitive load. Participants played a first-person target shooting game with either a virtual reality headset or on a monitor, using visual, visual-auditory, and a visual-haptic interfaces. Performance and self-reported data about cognitive load and spatial presence were collected for each interface. The study found that both immersive environments and multimodal interfaces reduced cognitive load. Additionally, an interaction was seen such that multimodal interfaces were more effective at reducing cognitive load in immersive environments. These findings indicate that the addition of more modalities into interfaces, especially in highly immersive environments, may be beneficial for decreasing cognitive load and increasing performance. Differences between types of multimodal interfaces emphasizes the need for future research into how different modalities contribute to cognitive load.

Acknowledgements

I extend my sincere gratitude to Ellie, Mom, Dad, Jarvis, Miranda, and Will for their unwavering support, occasional nudges, humor, love, encouragement, and wisdom throughout this journey. I could not have gotten to where I am today without your support. Your collective presence has added levity to the serious moments and provided steady anchors amidst my long academic journey. Thank you all. And thank you to Andrew, for being a reason to endure through it all.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Virtual Reality

In recent years, the use of virtual reality (VR) has seen widespread adoption as a standard learning and training tool, driven by the increasing availability and affordability of the necessary technology (Zawacki-Richter & Latchem, 2018). Broadly defined, VR refers to any system attempting to create a 3D environment for user interaction. Colloquially, VR is often associated with immersive virtual reality (IVR), a specific type of virtual reality system that involves the use of a headset (Robertson *et al.*, 1993). IVR has found applications across various industrial contexts, including medical, manufacturing, and corporate sectors. For instance, immersive virtual reality (IVR) has demonstrated remarkable effectiveness in training assembly line workers, resulting in an impressive 80% increase in post-training job performance compared to conventional methods (Gallegos-Nieto *et al.*, 2016). In medical training, IVR is employed to simulate realistic surgical scenarios, aiding in the preparation of surgeons for real-life conditions (Frederiksen *et al.*, 2020). Additionally, for scenarios requiring soft skills, immersive, interactive virtual reality allows trainees to experience realistic situations that respond to their actions in real-time, a role traditionally fulfilled by human actors (Li *et al.*, 2019). Despite being a relatively recent development, IVR builds upon the longstanding use of virtual reality in learning and training materials.

Popular 3D games such as *Grand Theft Auto V* and *Minecraft* are considered virtual reality due to their simulation of a 3D virtual environment. However, because these games are typically played on a TV or desktop monitor and not with a headset, they are known as nonimmersive virtual reality (NIVR). IVR is a system that attempts to create a more immersive 3D environment by having the user don a headset. These headsets offer the user six degrees of freedom in the visual exploration of the virtual environment, by allowing them to translate and rotate in three dimensions in their view of the virtual world by physically moving and rotating their head. These immersive systems may include peripheral accessories for users that enhance the immersion and intractability of the world, such as gloves that provide haptic feedback and track the movement of the user's hands or headphones that provide directional audio. These systems may also match the user's proprioceptive sense of body movements to the information that the user receives regarding their position in virtual space.

1.2 Immersion and Presence

The goal of many of these IVR systems is to create a heightened sense of presence that cannot be achieved in NIVR. Presence is the subjective feeling of being in a virtual environment, similar to how a person feels present in the real world. In the context of IVR systems, presence typically refers to the specific feeling of being in a virtual *spatial* environment (Cummings & Bailenson, 2015). Immersion and presence are closely related concepts; however, unlike presence, immersion is a quality of the system and the virtual environment that it produces. Immersion is independent of the user's perception and experiences, and it is a measure of the system itself, while presence is a subjective state that is self-reported by users. While different measures, immersion and presence are highly related to one another, as the immersion of a system often impacts the presence experienced by its users (Hou *et al.*, 2012).

The level of immersion provided by a system can impact the effectiveness of training as measured by the amount of material learned. Both IVR and NIVR can be useful for training and share some of the same characteristics (Makransky *et al.*, 2019). According to Slater and Wilber, a system’s level of immersion can be determined by examining the extent to which a system blocks out the physical world and how vividly it stimulates sensory experiences of its virtual environment (1997). Four dimensions of immersiveness emerged from this model: inclusivity, extensiveness, field-of-view (FOV), and vividity.

Inclusivity is the measure of how well a system shuts out reality. Through NIVR and IVR systems may both include accessories like noise canceling headphones, the headsets of IVR are unique in their ability to prevent users from seeing the physical world around them. Physical environments that do not distract from the virtual experience, such as a large room without physical distractions, also contribute to inclusivity. IVR that allows the user to explore a virtual environment by physically walking also uniquely benefits from physically large open spaces without obstacles as to allow users to freely move around without needing to consider real-world boundaries while in the virtual environment (Hashemian *et al.*, 2023).

Extensiveness is the measure of how many sensory experiences a system accommodates (Slater & Wilber, 1997). Both IVR and NIVR may increase their extensiveness by adding additional sensory experiences like audio and haptics. This can also extend to sensory experiences that are uncommon in virtual simulations, like temperature modulation (Ranasinghe *et al.*, 2017). The number of senses that a system simulates increases how immersive the system is, but visual simulation appears to contribute to a system’s immersiveness more than other senses (Slater & Wilber, 1997).

Because IVR and NIVR primarily rely on visual experiences, the visual field of view (FOV) of the system is considered its own dimension of immersion (Slater & Wilber, 1997). Both IVR and NIVR become more immersive when a wider visual FOV of the environment is displayed to the user, either by increasing the screen size of the computer monitor or the headset’s screen. Computer monitors can offer a wide range of FOVs, ranging from 10 to 90 degrees, while VR headsets typically offer FOV of around 90 to 110 degrees (Hou *et al.*, 2012; Infinite, n.d.). For reference, the human eye has a field of view of about 110-degrees horizontally while looking forward, meaning that IVR systems can approach near equivalency of FOV in the real world (Strasburger, 2020).

Finally, vividity is the measure of how thoroughly a system simulates a particular sensory experience. This aspect of immersion is perhaps the most diverse and difficult to identify due to the vast ways in which systems could differ in their emulation of a sensory experience. While some systems may aim to be indistinguishable from reality, others may be more stylized and intentionally distinct from reality; the realism of a simulation is not necessarily equivalent with how vivid it is, so the goals and design of the system must be considered when evaluating vividity.

The immersive attributes of a system do not all affect the sense of presence equally (Cummings and Bailenson, 2015). For instance, the refresh rate of the device both contribute greatly to the sense of presence experienced by users, while other details such as the pixel resolution of the images on the screen have a small effect on presence as seen in Table 1.1. This table shows that even NIVR systems can be considered more immersive if they contains other attributes that an IVR system lacks. Another interesting quality of the factors that contribute most to immersion is the bias towards visual information rather than other sensory information, like audio. This may indicate that a user’s sense of presence is derived from spatial visual cues, visual information used to orient oneself in an environment, rather than environmental realism as a whole (Cummings & Bailenson, 2015).

Table 1.1 Impact of Immersion Factors on Sense of Presence. The Pearson's r indicate the strength of the relationship observed by Cummings and Bailenson (2015)

Low-Impact ($r < 0.3$)	Medium-Impact ($r < 0.5$)	High-Impact ($r > 0.5$)
Image Quality	Field of View	Screen Framerate
Sound	Stereoscopy	
First Person View	IVR vs NIVR	

1.3 Immersive Virtual Reality and Cognitive Load

Despite immersive systems' ability to create a sense of presence, IVR's use in educational settings is contentious due to learners often becoming distracted and overwhelmed (Frederiksen *et al.*, 2020). Attributes that enhance presence in IVR and contribute to improved training outcomes also lead to higher cognitive load for users in comparison to those in NIVR. Users in IVR report experiencing both increased cognitive load and higher presence compared to NIVR (Makransky *et al.*, 2019). In an immersive environment, particularly those that utilize a VR headset to increase immersion, participants are flooded with more visual information; this information requires cognitive effort to process (Albus *et al.*, 2011). Therefore, some of the attributes of immersive systems that increase presence and contribute to positive learning outcomes in IVR may also decrease the actual learning of the participants. Despite increasing presence, IVR has been found to simultaneously reduce meaningful learning outcomes (Makransky *et al.*, 2019).

This observation is in accordance with cognitive load theory, a framework proposed by Sweller (1988). Cognitive load theory suggests that the brain can process a limited amount of information at a given time. As the brain engages in a particular task, it dedicates a portion of its limited resources to process the associated information. If current tasks demand too many cognitive resources, cognitive load will become too great and performance degradation will occur as resources are shifted from less highly prioritized tasks to more highly prioritized ones (Sweller, 1994).

Cognitive load is understood to be made up of three components: intrinsic, extraneous, and germane load (Sweller, 1994). The attributes of a task, the environment within which it is performed, and the characteristics of the person who performs it can all affect the amount and type of cognitive load experienced by learners. Intrinsic cognitive load, representing the innate difficulty of a task, is gauged through the concept of element interactivity, defined as the interconnectedness of task schemas. Sweller asserts that mastering interwoven schemas concurrently defines the difficulty of a task. In the realm of virtual environments, tasks exhibit a comparable intrinsic load to their non-virtual counterparts; for instance, a sophisticated medical simulation imposes a substantial intrinsic load akin to an actual medical procedure (Frederiksen *et al.*, 2020). Crucially, the immersive nature of a system remains inconsequential to intrinsic load; this measure stays tethered to the complexity of the task itself rather than the surrounding environment. Consequently, intrinsic cognitive load serves as a metric for evaluating the challenge inherent in learning a task, hinging on the intricate interconnectedness of the essential schemas (Sweller, 1994).

In contrast, extraneous cognitive load is contingent upon the medium and task presentation. Within Sweller's framework, it is shaped by both the presentation of information and the structure of the learning environment. The design of instructional materials should prioritize minimizing extraneous load to optimize learning outcomes. This load arises from non-essential elements introduced by instructional design choices. For instance, poorly designed diagrams or confusing instructions can heighten extraneous load, posing challenges for learners.

In virtual environments, the extraneous load may fluctuate based on interface design or the clarity of presented information. Unlike intrinsic load, extraneous load is significantly influenced by the immersiveness of the system. A system demanding the processing of excessive, irrelevant information would incur a high extraneous load on the user. Consequently, reducing extraneous cognitive load is paramount for effective learning, directly influencing learners' ease in comprehending and performing tasks (Sweller, 1994).

While intrinsic load relates the inherent difficulty of the task itself, there are additional elements that are non-essential to the actual task and are instead a product of task presentation. Irrelevant information and distracting environments must be processed by the brain, but do not contribute to the learning of the main task itself. For instance, the intrinsic cognitive load of a task consisting of solving an algebra equation and reading the text on a page is attributable to understanding, applying, and combining mathematical principles. However, the extraneous cognitive load would come from things such as redundant information, unintuitive or confusing presentations of the problem, or difficult-to-read text. In tasks that produce a low intrinsic cognitive load, the presence of some extraneous load due to these distractions may not significantly impact learnability or performance, due to the overall cognitive load being quite low (Sweller, 1994). However, as the intrinsic cognitive load increases due to task difficulty, the consequence of having extraneous cognitive load becomes more problematic. Therefore, reduction of extraneous load is of great concern in efforts to enhance learning environments (Sweller, 1994; Albus *et al.*, 2021).

While it is generally recommended that instructors reduce intrinsic and extraneous load, germane cognitive load is actually beneficial to learning environments. Germane cognitive load refers to the process of connecting learned material to preexisting knowledge to update or create new schemas. When new information is encountered, it can either be integrated into existing schemas through a process known as accommodation, or new schemas can be created or modified, through assimilation. Germane cognitive load, then, is the amount of effort needed by the mind to integrate that experience into the existing schemas, or to accommodate them by building entirely new schemas. Accommodation also produces a higher germane cognitive load than assimilation does. The creation or restructuring of schemas is more effortful than simply modifying existing ones. Creating and modifying schemas are closely associated with the learning process, as this allows learners to truly integrate the knowledge that they are learning and associate it with other knowledge that they already possess. These schemas are able to be stored in long term memory and can hold complex information while being understood as one construct (Kirschner, 2002). This compression and integration of new knowledge allows for the understanding of new information linked to the original schema as well as reducing cognitive load on future tasks. For instance, imagine a person new to immersive virtual reality (IVR) experiencing a complex training simulation. Initially, the user may feel overwhelmed as they navigate the virtual environment, manipulate objects, and engage with various elements. This cognitive load arises from the need to process a multitude of tasks simultaneously, such as spatial navigation, interaction with virtual objects, and understanding the rules of the simulated environment. However, as the user gains experience in the IVR setting, a cognitive schema specific to virtual reality interactions begins to form. Once this IVR schema is established, the user can approach new virtual environments or tasks with a more streamlined cognitive process. The once complex and overwhelming set of tasks now becomes a cohesive and manageable process, reflecting the assimilation of IVR-related information into their existing knowledge.

It may be that higher cognitive load experienced by those in IVR compared to those in NIVR is attributable to the greater amount of visual information that increases extraneous load (Tang *et al.*, 2022). In visual search tasks, for example, the presence of more visual noise increases task load (Chiossi *et al.*, 2023). These visual search tasks were paired with non-essential visual information, impacting performance on the primary task. This decrease in performance aligns with the predictions of cognitive load theory, wherein the processing of non-essential information adds to extraneous load, contributing to an overall higher workload (Sweller, 1988). Other research revealed that immersive VR training leads to poorer performance on various simulator metrics compared to conventional VR (Frederiksen *et al.*, 2020). Frederkiksen *et al.*'s study revealed that while IVR offered advantages by providing more realistic conditions, initial training in conventional VR was recommended before introducing IVR in surgical skills training due to the high cognitive load placed on trainees. In this scenario, it would appear that the extraneous cognitive load had less of an impact on performance once the germane cognitive load had lessened, due to the task load already being learned.

This could be explained by germane load (Sweller, 1994). Sweller argues that maximizing germane cognitive load should be the primary goal of instructional materials, as germane load involves the mental effort devoted to constructing task-related schemas, essential for effective learning. In the case of immersive virtual reality (IVR) training, where the environment is extraneously demanding due to the abundance of visual information, it becomes imperative to optimize germane load. Germane load plays a crucial role in acquiring and integrating essential information, fostering a deeper understanding of complex tasks. Given the limited capacity of the human mind for information processing, reducing extraneous cognitive load is crucial in IVR environments. The extraneous demands of IVR, marked by the presence of increased visual information, necessitate efficient cognitive resource allocation. Therefore, minimizing extraneous load is a strategic approach to ensure that learners can focus on and prioritize germane load. This approach becomes especially vital in environments like IVR, where the immersive experiences demand significant cognitive resources. By strategically managing extraneous cognitive load, instructional design in IVR can increase available cognitiveresources and allow for greater germane cognitive load, ultimately contributing to more effective learning outcomes.

IVR environments are especially prone to inadvertently increasing extraneous cognitive load. This is exacerbated by visually striking, rich environments featured in many IVR environments that distract users from the main task (Frederkson *et al.*, 2020). For instance, in a warehouse training simulation that uses virtual reality, a rich environment would include detailed interiors, real-time lighting, photorealistic textures, and plausible simulated physics. However, due to the computing and design costs of creating a rich virtual environment, training programs may decide to use a less realistic environment. Instead, representative or abstract environments can be used for non-essential and specific elements in the training simulation, creating a minimalist, non-immersive simulation. The benefits of these abstractions are twofold; not only do they reduce the processing and time capabilities for the simulation, but they also reduce the extraneous cognitive load by reducing unnecessary richness in the environment. As shown in the meta-analysis conducted by Cummings and Bailenson, image resolution and visual detail contribute very little to the immersiveness of a system (2015). Therefore, there is support for the idea that virtual environments designed for the purpose of learning should minimize distractions and unnecessary visual complexity to reduce extrinsic cognitive load.

1.4 Cognitive Load and Learning

However, there is research to suggest that environments designed for learning benefit from having extraneous information in the form of cues that direct the learner’s attention (Xie *et al.*, 2017). The integration of cues in multimedia learning materials addresses the challenge of high element interactivity, where learners may struggle to swiftly navigate and process relevant information within a constrained time frame. The addition of non-content information, as demonstrated in various studies, aims to guide learners' attention and alleviate the total cognitive load they bear. These additional cues offer information in redundant modalities, with the goal of increasing the number of ways a participant can engage with the learning material. For example, a lecture could be presented verbally by an instructor, while a transcript is given to students for them to read alongside the lecture. These redundant tasks or cues allow participants to use one or both of the modalities. While this theory proposes that these redundant cues should prevent cognitive overload, the empirical evidence remains inconsistent. The effects of cueing on learning outcomes aligns with the notion that cues may be related to an increased sense of presence, potentially explaining the observed improvements in learning performance (Chang *et al.*, 2017).

However, presence is known to increase as cognitive load increases, suggesting that not all elevations of cognitive load lead to negative effects in performance (Witmer, 1993). Recent studies have explored the possibility of dynamically adjusting task complexity in the context of virtual environments. By decreasing the visual complexity of a task in virtual reality based on the cognitive load experienced by an individual, task performance was able to be increased. In effect, the cognitive load of a task can be reduced by reducing the non-essential sensory components of that task (Chiossi *et al.*, 2023).

Strides have been taken to reduce cognitive load and increase learning in other virtual multimedia mediums. In these virtual learning systems, research has shown that presenting visual images alongside audible language provides the best environment for learning through the multimedia principle. (Mayer, 2005). The multimedia principle originates from the cognitive theory of multimedia learning, a theory that posits that the brain has different “channels” for processing visual spatial information and auditory verbal information, and that these channels each have a limited resource capacity. The brain can simultaneously attend to two components of a task if there is less interference between the tasks, a phenomenon originally noted by Wickens in Multiple Resource Theory (MRT; 2008). Traditionally, the modalities have been defined as being either visual or auditory but have since expanded to differentiate the focal and peripheral components of the visual modality, in addition to including the haptic modality alongside visual and auditory (Flemisch *et al.*, 2014). In addition to these modalities, the brain also processes spatial and verbal types of information (known as codes) somewhat separately. Figure 1.1 illustrates the separate processing of these types of information, all of which utilize different perceptual resources. However, they each incur a general cognitive load due to the sharing of some general shared resources. Because of this, interference of some degree will always occur when tasks are completed simultaneously; this interference will be greater when the modality and codes of the tasks overlap (Wickens, 2008).

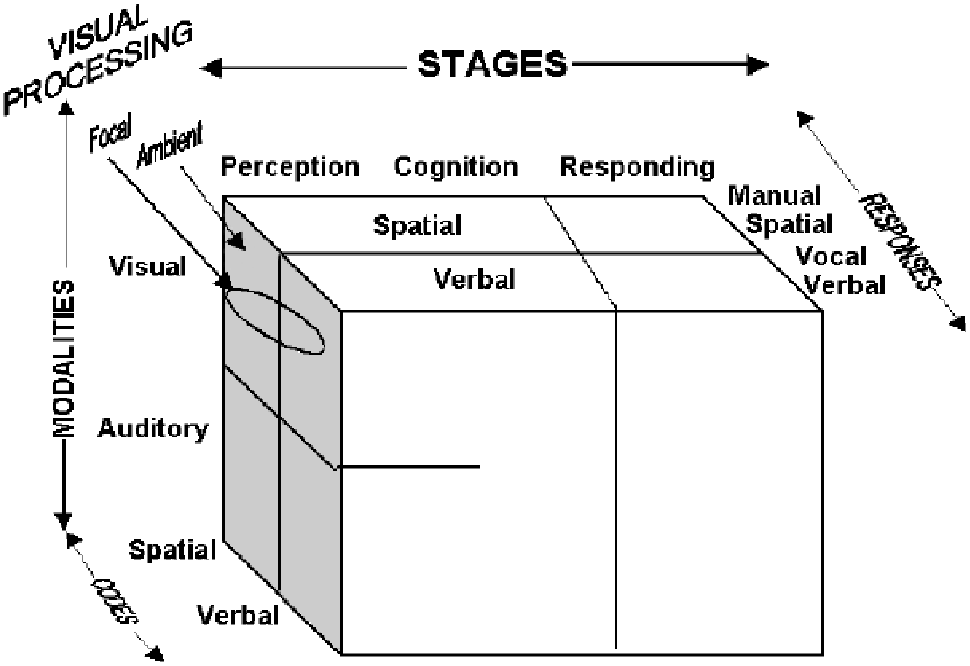
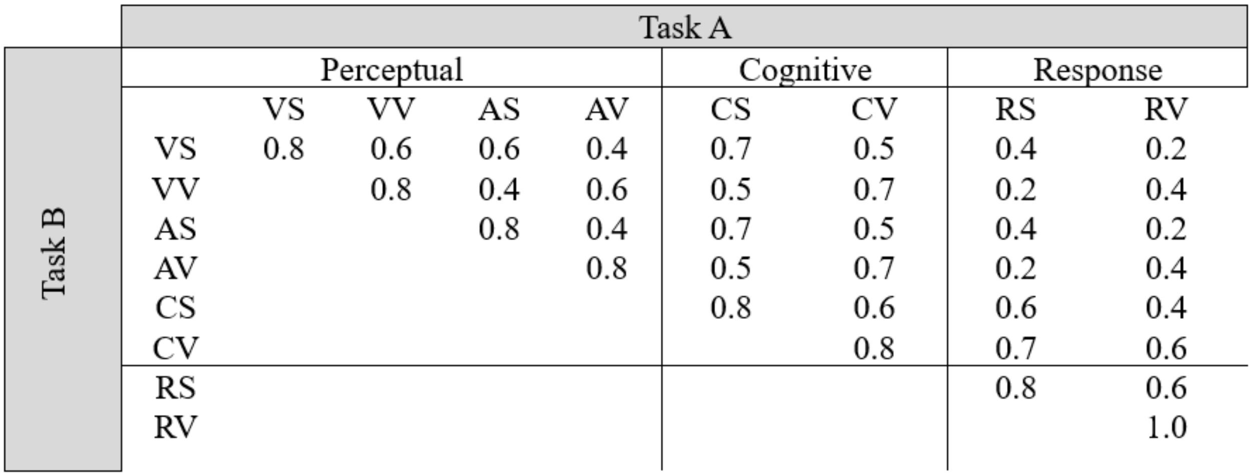


Figure 1.1 Multiple Resource Theory’s Model of Attention (Wickens, 2008).

Understanding the intricate relationships between different types of tasks and individual variations is crucial in predicting performance outcomes (Wickens, 2002). Task pairs that require a greater number of shared cognitive resources tend to produce greater degradation in performance, a phenomenon quantified through a conflict matrix modeling resource interaction (see Figure 1.2). This matrix assigns a conflict value ranging from zero to one, indicating the level of interference between tasks. A value of zero implies tasks can be performed simultaneously without interference, while a value of one suggests that the tasks cannot be performed simultaneously due to interference. Although achieving a conflict value of zero is unattainable due to the “cost of concurrence,” representing the minimal interference between two tasks sharing general cognitive resources, a low conflict value remains ideal for maximizing performance. For instance, the theory predicts that playing a first-person shooting game in IVR (a perceptual visual-spatial task) while simultaneously watching for a visual indicators on screen (another perceptual visual-spatial task) would yield a conflict value of 0.8. These tasks utilize the same channels with high interference, as both tasks engage the visual modality. In contrast, playing a first-person shooting in a virtual room while listening for an audio indicator (a perceptual audio-spatial task) would result in a conflict value of 0.6 because of the more differentiated processing.



		Task A							
		Perceptual				Cognitive		Response	
		VS	VV	AS	AV	CS	CV	RS	RV
Task B	VS	0.8	0.6	0.6	0.4	0.7	0.5	0.4	0.2
	VV		0.8	0.4	0.6	0.5	0.7	0.2	0.4
	AS			0.8	0.4	0.7	0.5	0.4	0.2
	AV				0.8	0.5	0.7	0.2	0.4
	CS					0.8	0.6	0.6	0.4
	CV						0.8	0.7	0.6
	RS							0.8	0.6
	RV								1.0

Figure 1.2 A conflict matrix, calculated for visual and auditory tasks of either spatial or verbal codes, for perceptual tasks, as noted by Multiple Resource Theory (Wickens, 2002).

Users perform better when interacting with systems that are designed to take advantage of multiple modalities as outlined by MRT. In a meta-analysis of automated driving, researchers found that automated vehicles that incorporated auditory or haptic feedback alongside visual representations performed better in task completion time than with interfaces that were solely visual (Weaver and DeLucia, 2022). Other research has found that interfaces that combined spatial and verbal cues induced a significantly lower cognitive load on users compared to interfaces with unimodal cues (Granados *et al.*, 2024). These principles apply to the design of virtual reality as well. In a recent study, interfaces with varying sensory modalities in virtual reality were compared to see how the presence of these modalities affected cognitive load and the sense of presence (Marucci *et al.*, 2021). The results from this study aligned with MRT's predictions: under situations with high task load, multisensory interfaces reduced cognitive load for users. Additionally, the study also found the condition with the most sensory modalities (visual, auditory, and haptic) produced the highest sense of presence in individuals. These findings provide evidence that systems designed to elicit a high sense of presence without increasing cognitive load would be ideal for creating a system that promotes learning and engagement. The use of additional sensory modalities in the training (haptic feedback) was more effective at increasing performance and decreasing reaction time than the systems that only utilized visual displays, and more complex tasks benefitted the most from immersive training compared to conventional training (Gallegos-Nieto *et al.*, 2016).

A few studies have examined Sweller's different types of cognitive load in immersive virtual reality tasks. It was found that immersive environments caused high extraneous and intrinsic cognitive load, but also a greater feeling of presence (Frederkson *et al.*, 2020). Makransky *et al.* (2019) also found that immersive virtual reality causes high extraneous and intrinsic cognitive loads as well as higher presence in a study comparing cognitive load between IVR and NIVR environments. As MRT would predict, there was an increase in cognitive load as more rich and visually complex processing was needed in the virtual environment and the resources used for visual-spatial processing may have been overwhelmed by the high amount of visual information. If this simulation had taken advantage of a multi-sensory system, extrinsic cognitive load may have been reduced. In doing so, the visual richness of the virtual world that contributes to higher presence would not have been affected. Additionally, while the authors categorized the experimental equipment into IVR and NIVR, the systems notably lacked control for various factors influencing immersion, including field of view, screen size, and degrees of freedom in movement. This omission introduces the possibility of confounding variables, especially concerning different types of controls and interaction methods. These factors play a crucial role in determining the user experience and could significantly influence cognitive load and learning outcomes. Recognizing this limitation emphasizes the necessity for more nuanced control in future research, aiming to gain a comprehensive understanding of the intricate relationship between immersion factors, cognitive load, and learning outcomes.

However, contrary to the efforts to reduce cognitive load, there is evidence that a high cognitive load may be linked to the beneficial sense of presence. Marucci *et al.* (2021) speculate that presence is modulated by multisensory systems alongside a high visual load, due to findings that indicated that users in systems with a high cognitive load reported greater feelings of presence. Achieving a sense of presence in immersive systems is a key strength, crucial for attaining a state of flow during a task. Flow, defined as a focused and enjoyable psychological state (Csikszentmihalyi, 1975), is characterized by concentration, filtering of irrelevant information, and pleasure—qualities beneficial for learning (Chang *et al.*, 2017). Flow has been conceptualized as arising from task concentration and intrinsic pleasure (Ghani *et al.*, 1991) as well as form control (meaningful interaction), attention focus (information filtering), curiosity (cognitive and sensory aspects), and interest (pleasure from the task; Webster *et al.*, 1993).

While efforts to minimize cognitive load in virtual environments have been a focal point, understanding the intricate interplay between cognitive load and the subjective experience of presence opens a new dimension in immersive learning experiences (Bjork & Bjork, 2011). A moderate amount of intrinsic load is ideal for training tasks and benefits the learning process. Increased cognitive effort is necessitated when individuals encounter challenging problems, prompting deeper processing and heightened engagement with the task at hand. Problems requiring retrieval or generation of information, rather than passive representation, contribute to more enduring and adaptable learning experiences, strengthening memory and understanding. The difficulty not only increases the intrinsic cognitive load, but also the sense of challenge and motivation. This motivational factor propels individuals to overcome obstacles, fostering a gratifying sense of accomplishment. These results may suggest that the more cognitively involved one is with a task, the more likely they are to feel present. This claim is also echoed in other studies that found a relationship between presence and an increase in cognitive load (Schrader and Bastiaens, 2012). These studies, however, do not distinguish the type of cognitive load being experienced by the users, so it is unclear which type was being affected by the addition of cognitive load.

The increase in presence observed in the high-immersion condition could be attributed to the richness of the visual media. However, the visual processing demands imposed by the high amount of visual information may overwhelm visual resources. One possible avenue for reducing cognitive load could have involved the implementation of a multi-sensory system within the virtual simulation. By distributing the cognitive load across multiple sensory modalities, the richness of the virtual environment could have been maintained while achieving a greater sense of presence and potentially leading users closer to a state of flow. Consequently, the intrinsic and extraneous cognitive loads could have been minimized.

1.5 Hypotheses

In the design of systems intended to train or educate people on how to perform a skill, balancing task difficulty, training time, and user ability is a challenge. A state of flow, or a state in which an individual is focused on and happy doing the task, is desirable because it improves learning. Achieving and maintaining that flow is essential to learning techniques. Previous literature has revealed that multimodal systems can influence the amount of cognitive load that users experience in virtual reality. The inclusion of more than one modality in an interface may decrease cognitive load and increase the sense of presence and flow. By designing an immersive reality system that incorporates principles of MRT, we may be able to lessen cognitive load to increase learning without dampening presence.

In this study, I draw upon MRT, Sweller's cognitive load theory, the relationship between presence and flow, and literature on redundant cues to investigate the impact of modality on cognitive load across difficulty levels, and how this relationship differs between NIVR and IVR systems. (Sweller, 1988; Wickens, 2008). I introduce three cue interfaces—visual, visual-auditory, and visual-haptic— that are used to inform decision making in a game, which allow for the testing of my hypotheses. In my experiment, participants will play through a game in either IVR or NIVR, with varying levels of difficulty. Participants will play through the game three times, experiencing each cue interface: a visual cue (visual condition), a combination of visual and auditory cues (visual-auditory condition), and a combination of visual and haptic cues (visual-haptic condition). The game will require participants to simultaneously perform two separate tasks; they must aim and shoot targets that appear for a limited time while keeping track of a gun’s readiness to shoot. The readiness of a target can be communicated through these visual, audio, or haptic cues.

According to the literature on redundant cueing, additional cues should alleviate cognitive load (Xie *et al.*, 2017). However, MRT and Sweller’s cognitive load theory suggest that the addition of more cues that the mind must process would lead to an increase in cognitive load (Sweller, 1988; Wickens, 2002). Therefore, I offer competing hypotheses.

H1 – Multimodal Interface Main Effect: Literature on redundant cueing predicts that participants will experience lower cognitive load when interacting with multimodal interfaces than with unimodal interfaces ($H1_a$), while MRT predicts that participants will experience increased cognitive load in multimodal interfaces ($H3_b$).

H2 – Immersion Main Effect: Flow theory predicts that participants will experience more presence and lower cognitive load in the immersive condition than in the non-immersive condition ($H2_a$), while MRT predicts that participants will experience increased cognitive load in the immersive condition than in the nonimmersive condition ($H2_b$).

H3 – Multimodal Interface and Difficulty Interaction: Literature on redundant cueing predicts that the lower cognitive load experienced in the multimodal interfaces will become more pronounced as difficulty increases, ($H3_a$), while MRT predicts that the higher cognitive load experienced in the multimodal interfaces will become more pronounced as difficulty increases, ($H3_b$),

H4 – Immersion and Difficulty Interaction: Flow theory predicts that the lower cognitive load experienced in the multimodal interfaces will become more pronounced as difficulty increases, ($H4_a$), while MRT predicts that the higher cognitive load experienced in the immersive environments interfaces will become more pronounced as difficulty increases, ($H4_b$).

H5 – Multimodal Interface and Immersion Interaction: Literature on redundant cueing and flow theory predict that the lower cognitive load experienced in the multimodal interfaces will become more pronounced in the immersive condition, ($H5_a$), while MRT predicts that the higher cognitive load experienced in the multimodal interfaces will become more pronounced in the immersive condition ($H5_b$).

H6 – Multimodal Interface, Immersion Level, and Difficulty Interaction: Flow theory and literature on redundant cueing predict an interaction between multimodal interfaces and immersive virtual reality that will become more pronounced as difficulty increases, leading to a more substantial reduction in cognitive load as difficulty increases ($H6_a$). MRT also predicts an interaction between multimodal interfaces and immersive virtual reality that will become more pronounced as difficulty increases, leading to a more substantial increase in cognitive load as difficulty increases ($H6_b$).

Chapter 2. Method

2.1 Participants

Participants (*N* = 59) were collected from SONA (*n* = 43), an experiment management and scheduling system. Participants were also recruited with fliers posted in Facebook groups pertaining to the university and at various physical locations on campus (*n* = 16). Students who signed up through SONA received course credit as compensation. Participants who responded to the fliers received a 2-liter soda and entry into a drawing for a $100 debit card as compensation. Participants who self-reported uncorrected vision or sensitivity to motion sickness were excluded from the study. The mean age of participants was 22.4 years old and 81% claimed to play video games at least once a week. This research complied with the American Psychological Association's Code of Ethics and was approved by the Institutional Review Board at the University of Alabama in Huntsville.

2.2 Materials

Monitor and Headset: The game was presented to participants using either a 32-inch display monitor or an HTC-Vive headset. The HTC-Vive headset displayed the game on two high-resolution OLED screens positioned in front of the eyes and held in place with a head strap. The HTC-Vive uses two base stations to track the motion of the user and the headset so that the user can be correctly positioned in the virtual world.

Controllers: All participants interacted with the game using a single HTC-Vive controller. These controllers offer precise positional tracking and various input options, including buttons, triggers, and a touchpad (see Figure 1.3). Additionally, the controllers can vibrate to provide feedback to the users. The positional tracking of the controllers allows for an on-screen cursor to be moved around the screen. Selecting an option or shooting a gun was performed by pressing the trigger on the controller.

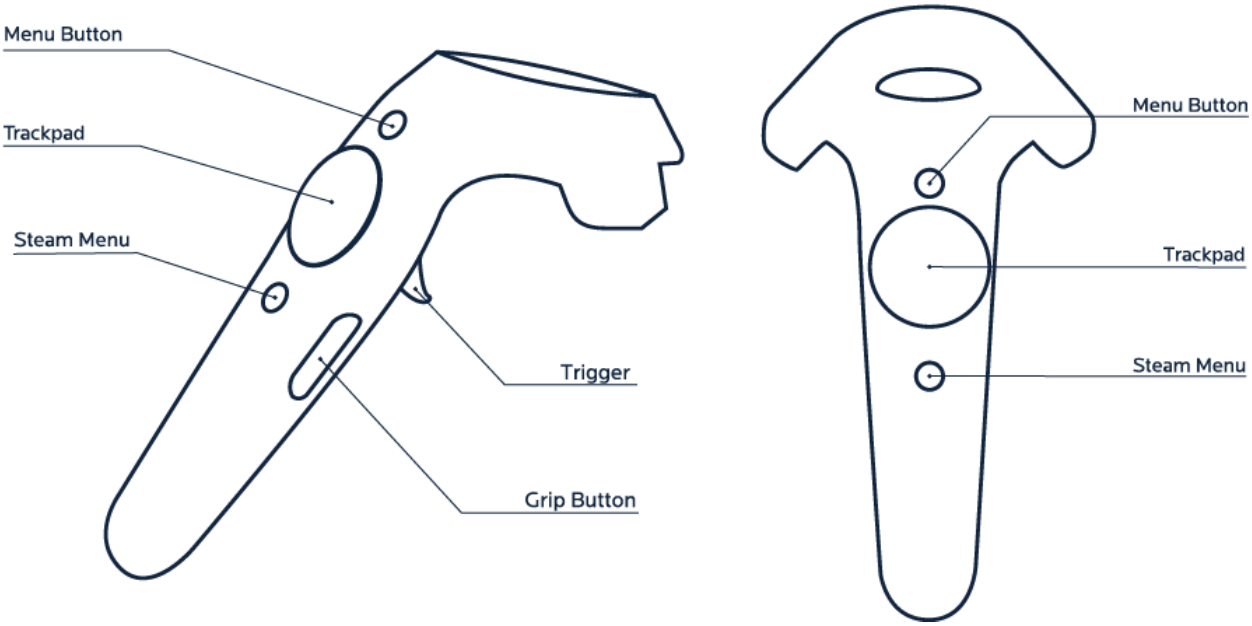


Figure 1.3 HTC-Vive controller. *From VR Games, apps, & videos*. VIVEPORT. (n.d.). https://www.viveport.com/

Demographics: Participants provided information about their age, gender, gaming experience, and virtual reality headset experience. This questionnaire also served to screen participants for normal or correct vision and for motion sickness sensitivity.

NASA Task Load Index: The NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988), is a widely-used, 6-item self-report instrument for assessing subjective workload and cognitive load. It employs a 21-point Likert scale to measure cognitive load across six dimensions: mental demand, physical demand, temporal demand, perceived success, effort, and frustration. A composite score was obtained by subtracting 1 from a participant’s response on each item and multiplying the result by 5. These items were then averaged together to create a composite workload score.

Multimodal Presence Scale: The Multimodal Presence Scale (MPS), specifically the Physical Presence subscale, were employed in this study to assess participants' perceived sense of spatial presence within the virtual environment. The MPS Physical Presence subscale consists of five items rated on a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7; Makransky *et al.*, 2017). Items included statements such as, “the virtual environment seemed real to me.” The presence composite score was calculated by averaging together the rating from all items.

2.3 Experimental Task

Cue Interfaces. Participants played through three blocks, one for each variation of the cue interface. The three cue interfaces were visual, visual-auditory, and visual-haptic. The order of these blocks are counterbalanced with a Latin square design. In all blocks, participants were shown the visual cue. The visual cue interface was a circular deep-red flash of color that appeared on the back of the virtual gun (see Figure 1.4). When the gun was ready to shoot, the flash appeared on the back of the gun for one second before disappearing. The gun was able to fire as soon as the cue appeared. In the visual-auditory condition, the visual cue was shown as described as above, but an auditory cue played concurrently with the visual cue. The auditory cue was a second long sound effect of a gun being reloaded. In the visual-haptic condition, the visual cue was also shown, but the controller vibrates alongside it. The vibration was a strong pulse that lasted one second. Before each block started, participants saw a short countdown and some text that informed them about the upcoming cue. The text read either “WATCH”, “WATCH & LISTEN”, or ‘WATCH AND FEEL” depending on whether the upcoming cue interface would be visual, visual-auditory, or visual-haptic, respectively.

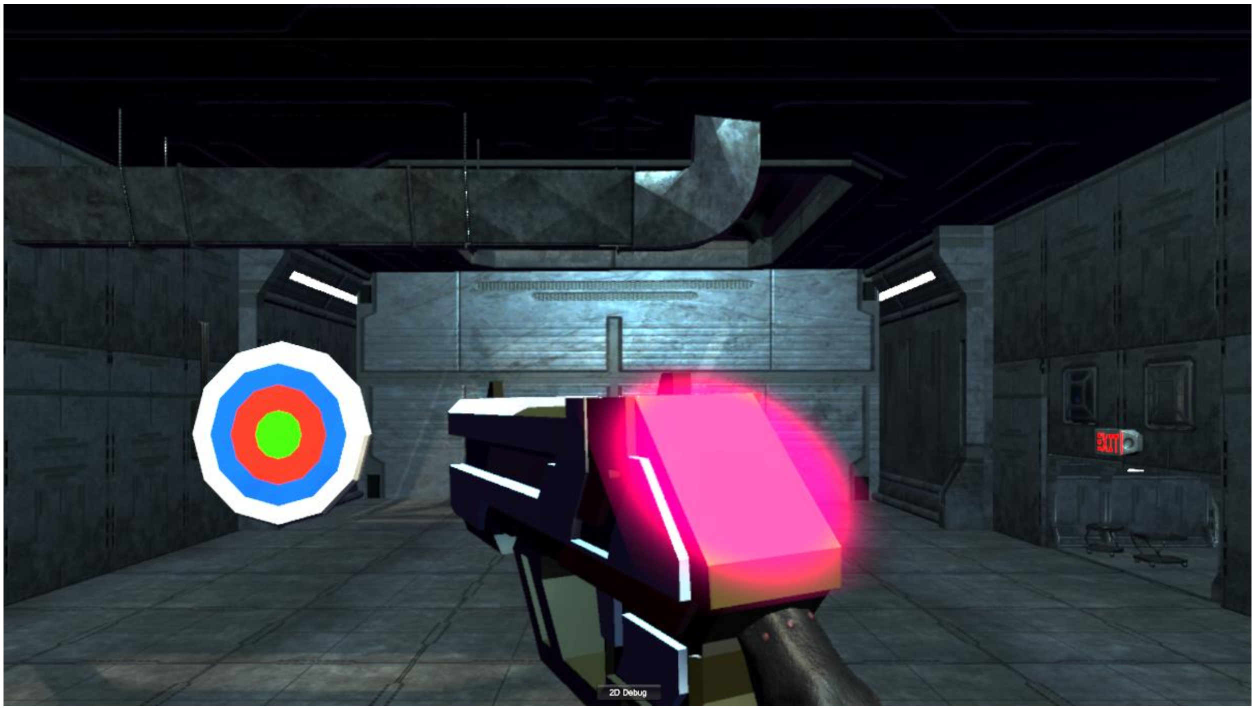


Figure 1.4 Screenshot of the visual cue present in all interfaces.

Immersion Levels. Participants were alternatively assigned to be in either the immersive or nonimmersive condition. If a participant was assigned to the immersive condition, they were given the HTC-Vive headset and used it to view the game. If a participant was assigned to the nonimmersive condition, they instead used the desktop monitor to view the game. Participants in both the immersive and nonimmersive conditions were given headphones and an HTC-Vive controller to use throughout the study.

Difficulty Levels. Difficulty was used in order to manipulate cognitive load, with higher difficulties used to induce higher levels of cognitive load. Difficulty was manipulated by adjusting the size and travel speed of the target. The target size could range from X to Y and was determined by defining three equidistant points on the x,y,z coordinate plane. Size and distance were measured in Unity units, a constant unity of measurement that is universal throughout the Unity program (Unity, 2023). Each point was defined by the function:

$$300 - (10 * \text{difficulty}),$$

where difficulty is a value that ranges from 1 to 15. The z-axis (how thick the targets were) was held constant at 10 throughout the study. Target speed was measured as units per second. The target speed could range from .256 to 3 and was defined by the function:

$$300 - (10 * \text{difficulty}).$$

Using the same difficulty value as the difficulty used to determine the size of the target, ranging from 1 to 15. This difficulty modifier changed every 12 targets throughout each block, for a total of 15 difficulty levels per block. 14 out of the 15 difficulty levels are shown to participants. The presentation order of the difficulty was randomized for each participant and consistent across blocks. Each difficulty level would only be shown once until all 14 difficulty levels had been shown, after which the levels begin to loop.

Tutorial Level. When starting the game, participants were placed into a tutorial level where they must shoot 5 targets before progressing onto the main game. This tutorial level took place in the same environment as the main game. In this level, no data were collected and no cues or moving targets were present. Instead, targets appeared in the corners of the screen and did not move. While in the tutorial level, the gun had no delay. Participants attempted to shoot these targets as many times as they wanted without the target moving. Each time one of these targets was shot, it disappeared and another reappeared in a different place. Participants were verbally instructed to practice using the gun in this area. Text at the bottom of the screen read “Shoot the Targets!” as a reminder. This text was only present for the tutorial level. After enough targets were shot, a green box appeared at the top of the level that had the text "Shoot me to Start" written on it. Participants were allowed to shoot as many targets as they wish before shooting the start button. After shooting the start button, the game immediately began, starting with whichever cue interface the participant had been assigned to first.

2.4 Pilot Testing

Participants (N = 5) were collected from a convenience sample based on their varied experience with video games and virtual reality. Four participants had experience with video games, and two also had experience with virtual reality. These participants played through prototype versions of the game to determine its optimal length and difficulty. The bounds for difficulty were determined based on this testing, with the upper limit of difficulty (15) being set to conditions in which participants successfully hit fewer than 50% of targets, and the lower limit of difficulty (1) set to conditions in which participants hit more than 90% of targets. The length of the game was based upon achieving a total gameplay session length of around 45 minutes.

2.5 Procedure

Participants completed the Qualtrics survey containing the consent form and demographic questionnaire, then received verbal instructions about how to play the game. A full transcription of the instructions is available in the appendix. Upon starting the game, participants completed the tutorial level that introduced them to the controls of the game and allowed them to become acquainted with shooting targets. Participants then played through a block featuring the first cue interface. After completion of the block, participants answered the NASA TLX and MPS questionnaires, and then started the next block. After answering the final set of questions, the game ended. The participants were then debriefed and given their compensation.

Chapter 3. Results

Data were analyzed using R in RStudio using the emmeans and lme4 packages (R Core Team, 2023; Leneth, 2024; Bates *et al.*, 2015). The results were visualized using the ggplot2 and effects packages (Wickham, 2016; Fox & Weisberg, 2009).

3. 1 Data Cleaning

Performance was used as a measure of cognitive load and was captured by whether participants hit (1) or missed (0) the target (due to repeated inaccuracies or misfires) on a given trial. Video game experience and VR experience were assessed by the number of hours participants self-reported playing on each type of system. Before analyses were conducted, the categorical variables, cue and immersion, were effect-coded while the difficulty, video game experience, and VR experience, the continuous variables, were means-centered.

3.2 Manipulation Checks

To ensure that immersive virtual reality increased users' sense of presence, as predicted by a higher level of immersion (Bjork & Bjork, 2011), a manipulation check was performed by conducting a repeated-measures regression to examine the effects of immersion condition on the self-reported sense of presence. This revealed that difficulty did negatively impact performance, $B = -0.25$, $SE = .08$, $z = -54.84$, $p < .001$, ensuring that increasing the difficulty manipulation of the game degraded performance as expected. A post-hoc Tukey's HSD revealed that the immersive condition did increase presence, $B = -1.19$, $SE = .22$, $\mathrm{t} = -3.17$, $p = .003$, relative to the non-immersive condition.

3.3 Hypothesis Tests with Behavioral Measures

According to Multiple Resource Theory, it was hypothesized that the combination of multimodal cues and immersion would result in higher cognitive load (Wickens, 2008). Specifically, an interaction effect was predicted, suggesting that the increased cognitive load due to multimodal cues would be more pronounced in the immersive condition relative to nonimmersive one. However, flow theory predicted that increased immersion would lead to increased presence and inducement of the flow state, decreasing cognitive load. Literature on redundant cueing suggested that the addition of multiple cues in an interface should reduce cognitive load. To test the predictions of my hypotheses, a repeated-measures logistic regression examined the effects of immersion, cue interface, and difficulty level on performance. Weekly hours of video games played and the presentation order of the targets per block were added to the model as covariates to improve model fit and to control for possible confounding variables.

Hypothesis 1. To determine differences in cognitive load between interfaces, a post-hoc Tukey's HSD was conducted to determine differences in performance between the different interfaces. As can been seen in Table 3.1 and Figure 3.1, the probability of a participant hitting a target on a given trial was increased in the visual-haptic interface as compared to the visual interface, while no significant differences were found between the visual and visual-auditory interfaces and the visual-auditory and visual-haptic interfaces.

Table 3.1 Pairwise Comparisons of Cue Interface Condition on Performance. Estimates represent the differences in performance across interface types.

Comparison	B	*SE*	*B*	*SE*	*z*	*p*
Visual-auditory to Visual Interface	0.05	.05	.07	.04	1.10	.51
Visual-haptic to Visual Interface	0.05	.05	.19	.04	2.83	.01
Auditory Visual to Visual-haptic Interface	-0.08	.05	-.12	.04	-1.69	.21

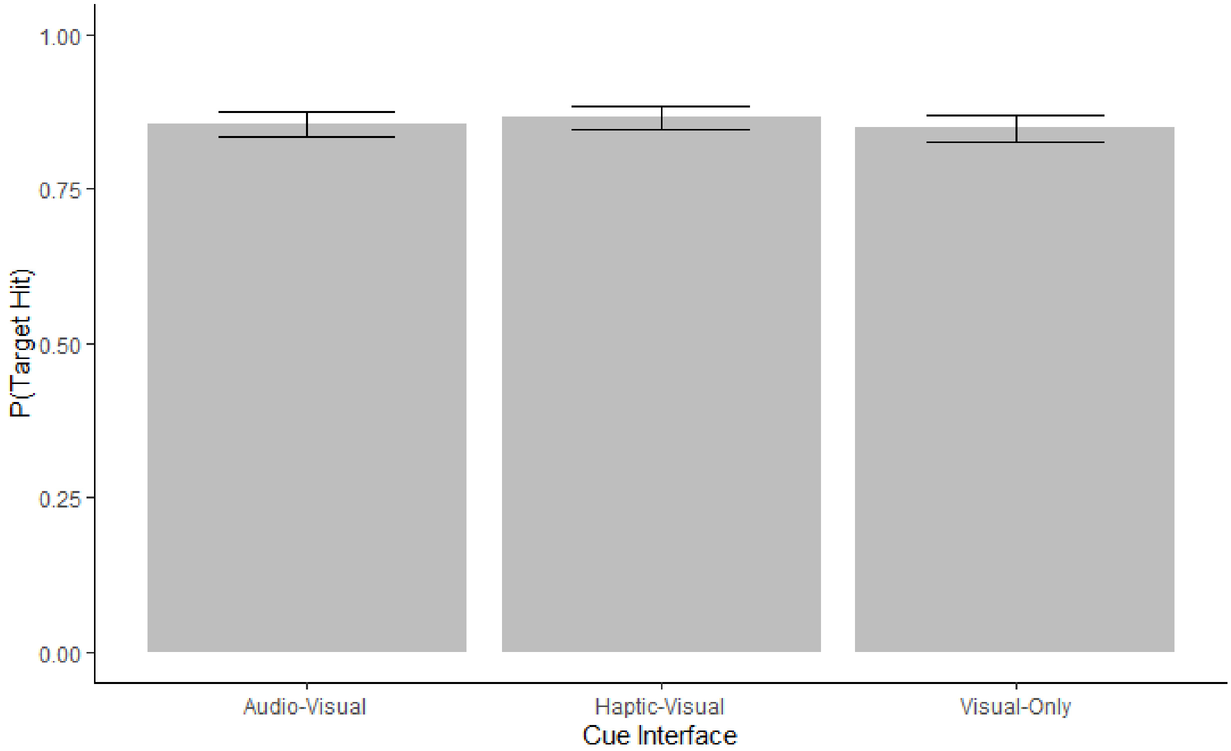


Figure 3.1 Performance as a Function of Cue Interface Condition. Error bars indicate 95% confidence intervals.

Hypothesis 2. To test the predictions of H2, a post-hoc Tukey's HSD was conducted to determine whether differences in performance differed between the immersion conditions. As seen in Figure 3.2, the probability of a participant hitting a target on a given trial was higher in the immersive condition compared to the nonimmersive condition, $B = .46$, $z = 2.75$ $SE = .12$, $p = .01$.

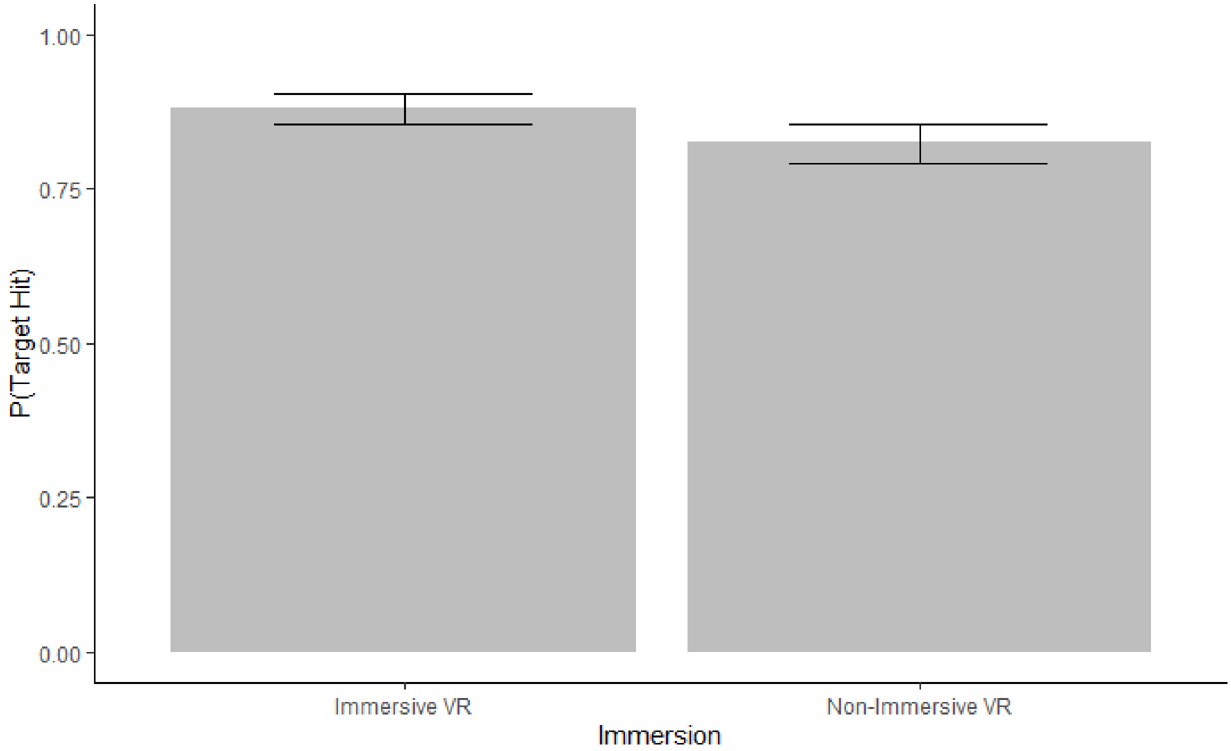


Figure 3.2 Performance as a Function of Immersion Condition. Error bars indicate 95% confidence interval.

Hypothesis 3. To test H3, another Tukey's HSD was conducted to determine whether the relationship between difficulty level and performance differed across the different cue interfaces. As can be seen in Figure 3.3 and Table 3.2, the probability of a participant hitting a target on a given trial was higher in the visual-haptic interface as compared to the visual-auditory interface, while no significant differences were found between the visual and visual-haptic interfaces and the visual and visual-auditory interfaces. These results fail to support $H3_a$, as there was no significant increase in performance across difficulty levels and a decrease with the visual-auditory interface compared to the visual interface. However, these results do not support $H3_b$ either due to the visual interface performing similarly to the visual-haptic interface.

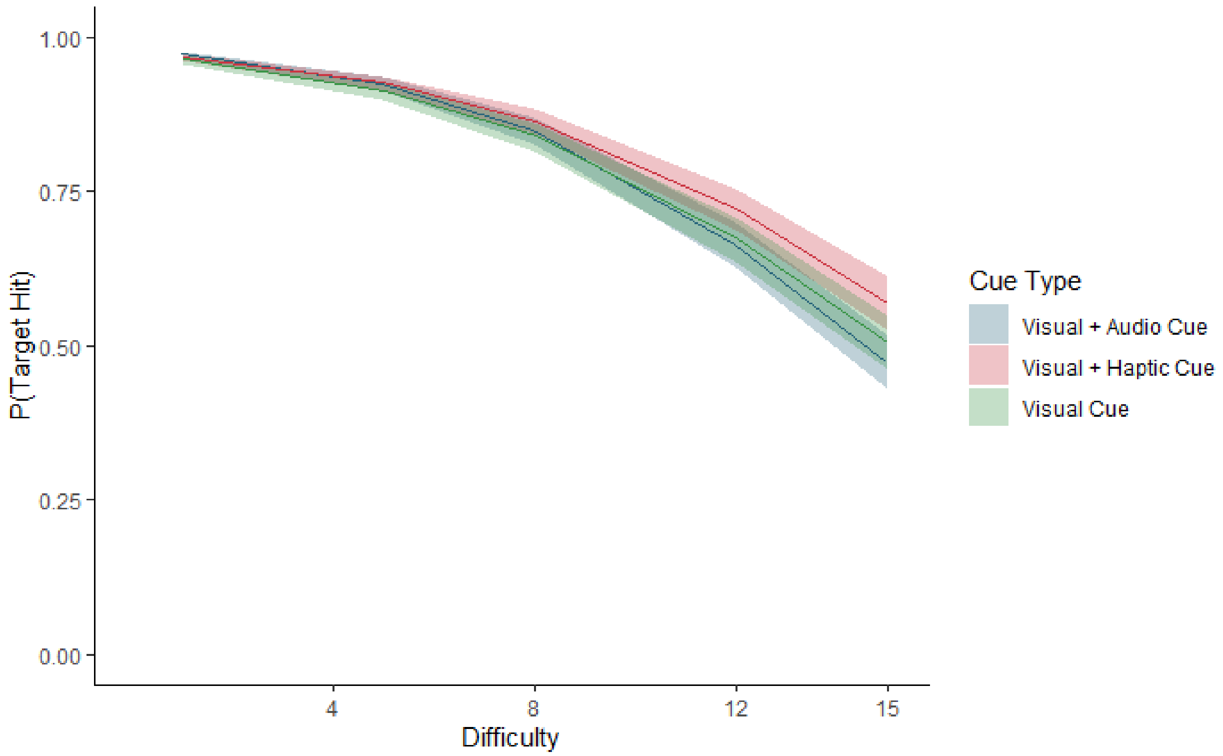


Figure 3.3 Performance as a Function of Difficulty Slope and Cue Interface Condition.

Table 3.2 Pairwise Comparisons of the Difficulty Slope between Cue Interface Interface Conditions. Estimates represent the differences in performance across interface types.

Comparison	B	*SE*	*B*	*SE*	*z*	*p*
Visual-auditory to Visual Interface	-.02	.01	-.02	.01	-2.17	0.08
Visual-haptic to Visual Interface	.02	.1 × 10^{-3}	.01	.1 × 10^{-3}	1.71	0.20
Auditory Visual to Visual-haptic Interface	-.04	.01	-.03	.01	-3.82	<.001

Hypothesis 4. To test the predictions of H4, a post-hoc Tukey's HSD was conducted to determine differences in performance between the different immersion levels across difficulty levels. As seen in Figure 3.4, the probability of a participant hitting a target on a given trial was higher in the immersive condition compared to the nonimmersive condition ($B = .02$, $SE = .1 \times 10^{-3}$, $z = 2.77$, $p = .01$). These results support $H4_a$, indicating that the effectiveness of immersion on cognitive load becomes more pronounced as difficulty increases.

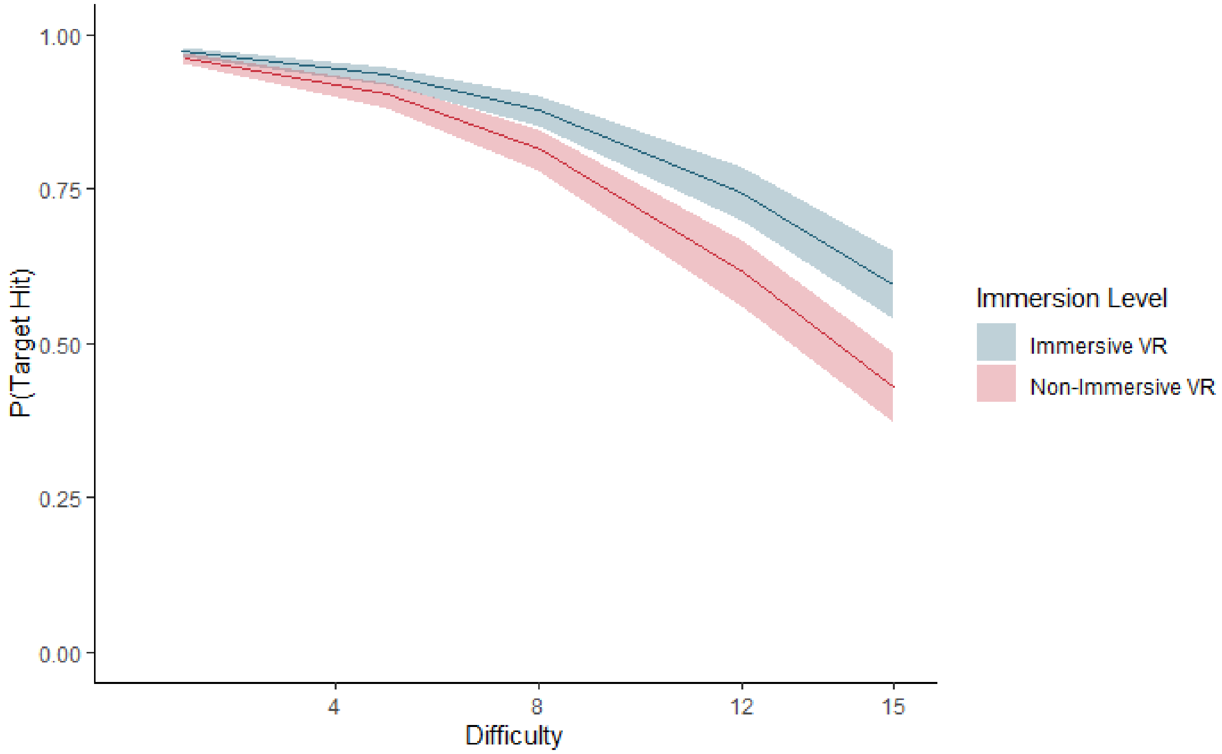


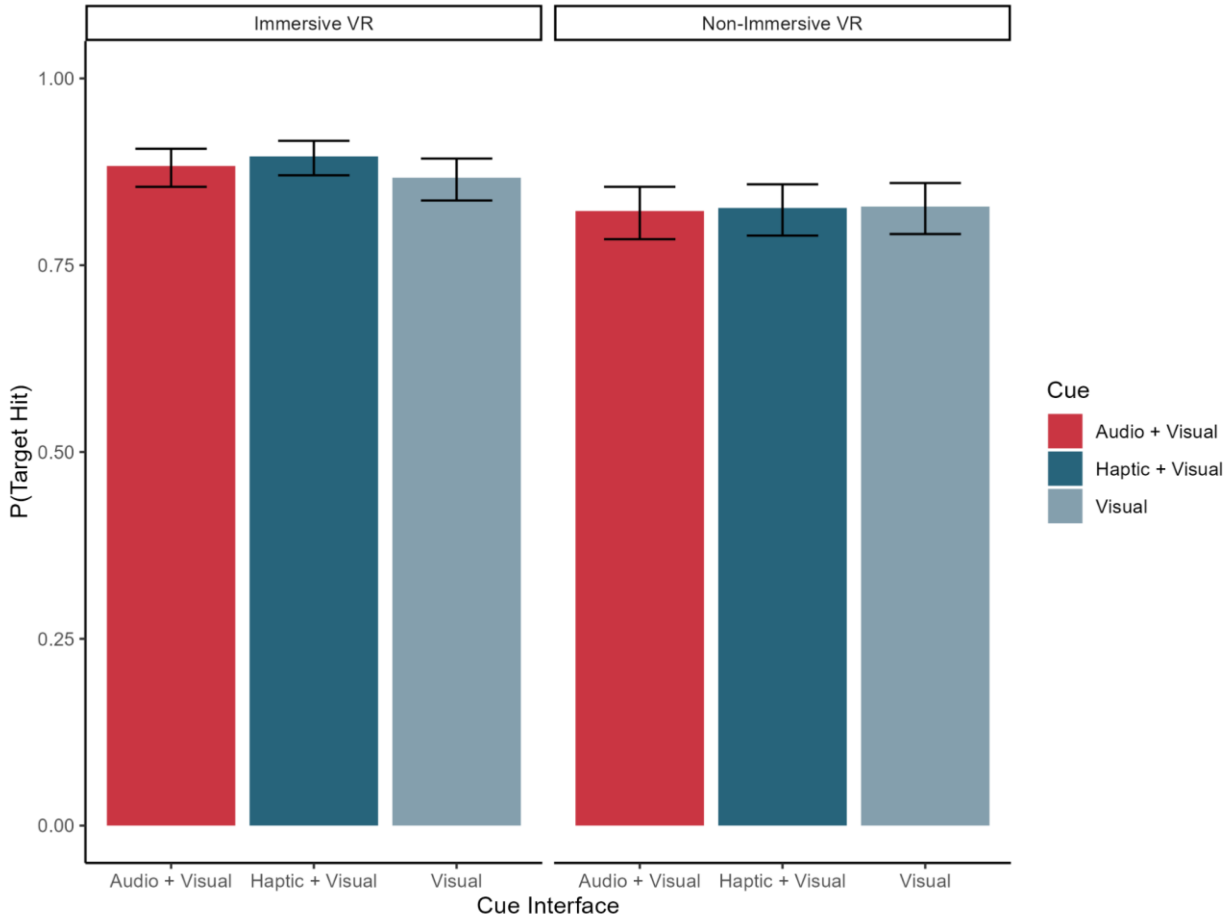
Figure 3.4 Difficulty Slope as a Function of Immersion Condition. Shaded regions bars indicate 95% confidence intervals.

Hypothesis 5. To test H5, another Tukey's HSD was conducted to determine the differences in performance between the different interfaces across immersion levels. In nonimmersive virtual reality, no significant differences existed between the visual and visual-auditory interfaces, the visual and hpatic-visual interfaces, and the visual-auditory and visual-haptic interfaces (see Table 3.4). However, in immersive virtual reality, the visual-haptic interface outperformed the visual interface, though there are no significant differences between the visual-auditory and visual interfaces and the visual-auditory and visual-haptic interfaces, as can be seen in Figure 3.5. These results support $H5_a$, suggesting that multimodal interfaces are more effective at reducing cognitive load in immersive environments.

Table 3.3 Pairwise Comparisons of Cue Interface and Immersion Conditions on Performance Note. Estimates represent the differences in performance across interface types and immersion levels.

Comparison	B	*SE*	*B*	*SE*	*z*	*p*
Immersive VR						
Visual-auditory to Visual Interface	-.04	.06	-.02	.06	-.65	.79
Visual-haptic to Visual Interface	-.01	.06	.05	.06	-.21	.98
Auditory Visual to Visual-haptic Interface	-.03	.06	-.07	.06	-.44	.90
Non-Immersive VR						
Visual-auditory to Visual Interface	.14	.07	-.17	.06	2.13	.08
Visual-haptic to Visual Interface	.27	.07	.34	.06	4.08	<.001
Auditory Visual to Visual-haptic Interface	-.13	.07	-.17	.07	-1.88	.15

Figure 3.5 Comparison of Performance by Cue Interface Condition between Immersion Condition. Error bars indicate 95% confidence interval.



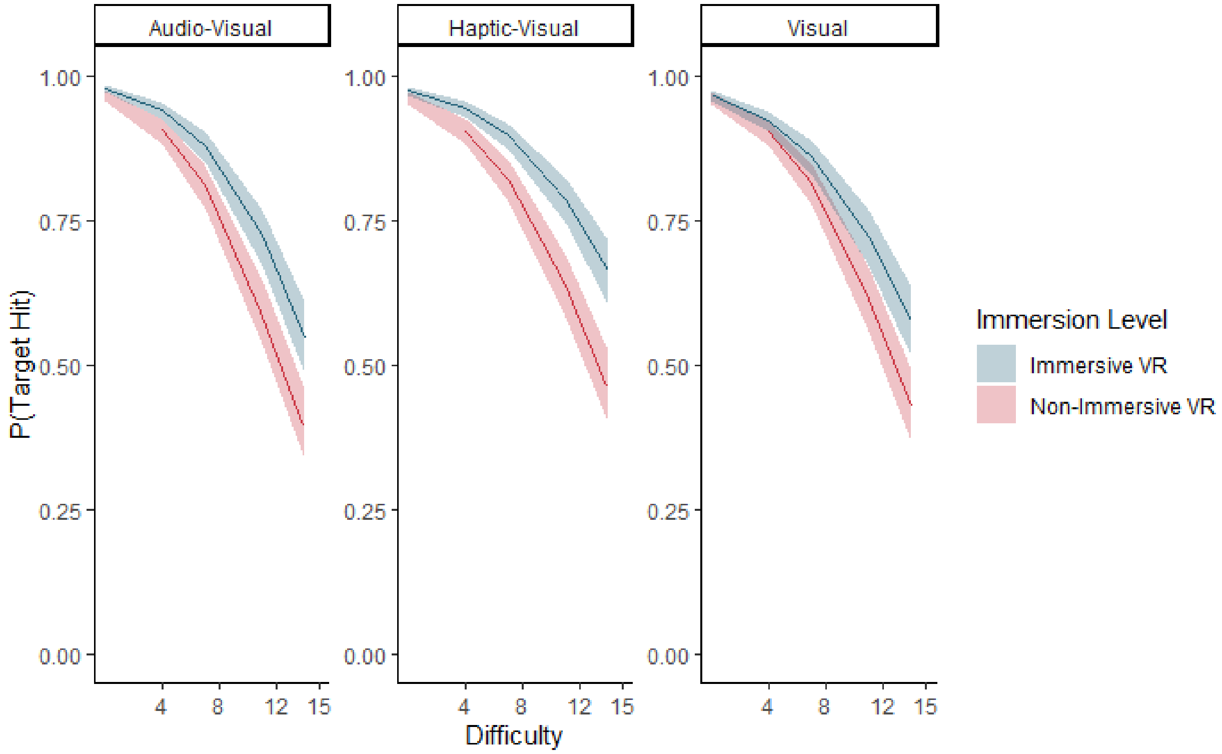
Hypothesis 6. Finally, to test H6, Tukey's HSD was conducted to determine the differences in performance between the different interfaces across immersion levels and across difficulty levels. As can been seen in Table 3.5, in non-immersive virtual reality, no significant differences existed between the visual and visual-auditory interfaces, the visual and visual-haptic interfaces, and the visual-auditory and visual-haptic interfaces across difficulty. However, in immersive virtual reality, the visual-auditory interface performed worse than the visual-haptic interface as difficulty increased, but no significant differences were found between the visual-haptic and visual interfaces and the visual-auditory and visual interfaces. These results indicate a

decrease in performance as difficulty increases while in immersive virtual reality, which can be seen in Figure 3.6. This supports $H6_a$ and suggests multimodal interfaces may induce higher cognitive load in immersive environments as difficulty increases, but it is important to note that the visual-haptic interface experienced no degradation in performance as difficulty increased compared to the visual interface.

Table 3.4 Pairwise Comparisons of Difficulty Slope across Cue Interface and Immersion Conditions. Estimates represent the differences in performance across interface types and immersion levels.

Comparison	B	*SE*	*B*	*SE*	*z*	*p*
Non-Immersive VR						
Visual-auditory to Visual Interface	-.04	.02	-.02	.01	-2.23	.07
Visual-haptic to Visual Interface	.02	.01	.01	.01	1.09	.52
Auditory Visual to Visual-haptic Interface	-.05	.01	-.03	.01	-3.21	.004
Immersive VR						
Visual-auditory to Visual Interface	-.01	.02	-.04	.01	-.82	.69
Visual-haptic to Visual Interface	.02	.02	-.05	.02	1.33	.38
Auditory Visual to Visual-haptic Interface	-.03	.02	-.01	.01	-2.16	.08

Figure 3.6 Performance as a Function of Immersion Condition, Cue Interface Condition, and Difficulty. Shaded regions indicate 95% confidence intervals.



3.4 Hypothesis Tests with Self Report Measures

Hypothesis 1. As an additional comparison of cognitive load, Tukey's HSD was conducted to determine the differences in subjective workload as measured by a participant's composite TLX score between the different interfaces (see Table 3.2). As can been seen in Figure 6, participants self-reported lower workload using the hapic-visual interface as compared to the visual interface, while no significant difference was found between the visual and visual-auditory interfaces and the visual-auditory and visual-haptic interfaces. The results from these analyses support $H1_a$, suggesting that multimodal interfaces can reduce cognitive load compared to unimodal interfaces.

Table 3.5 Pairwise Comparisons of Cue Interface Condition on TLX Score. Note. Estimates represent the differences in performance across interface types.

Comparison	B	*SE*	*B*	*SE*
Visual-auditory to Visual Interface	-2.33	1.28	-1.76	.16
Visual-haptic to Visual Interface	-4.60	1.28	-3.58	.001
Auditory Visual to Visual-haptic Interface	2.27	1.28	1.76	.18

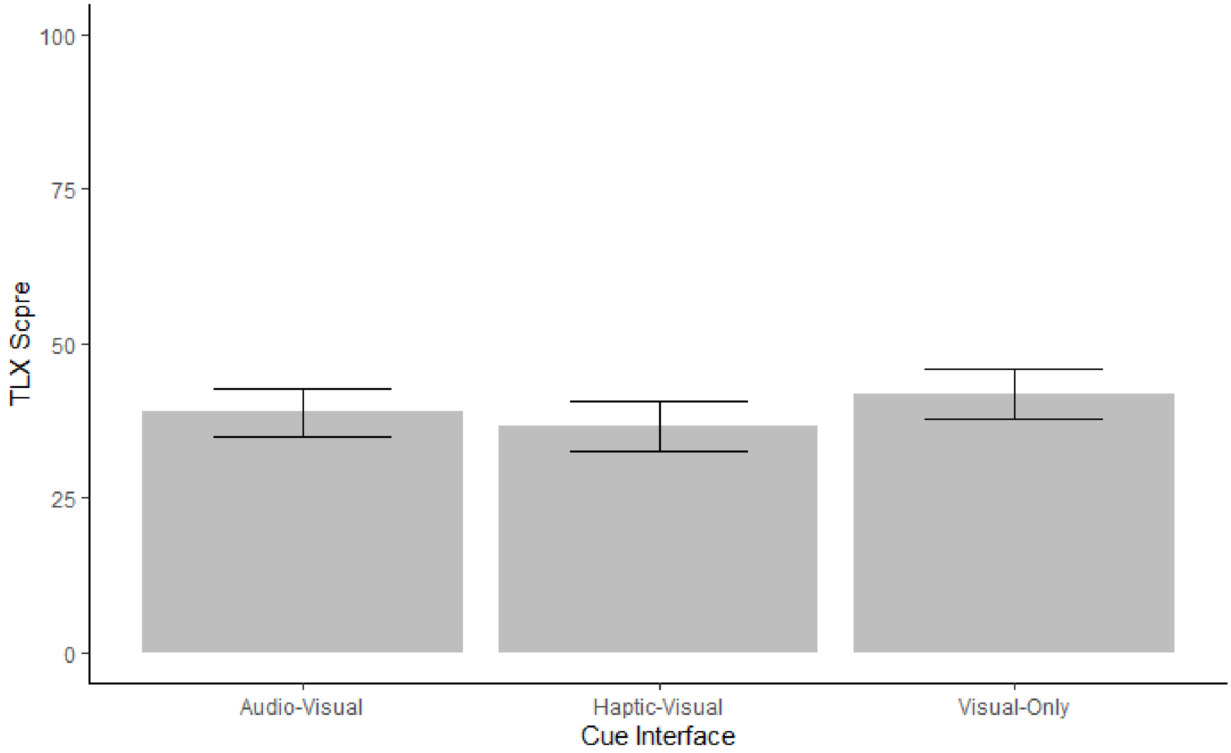


Figure 3.7 TLX Score as a Function of Cue Interface Condition. Error bars indicate 95% confidence interval.

Hypothesis 2. An additional Tukey's HSD was conducted to determine the differences in TLX scores between the levels of immersion, but no significant differences were found between the groups.

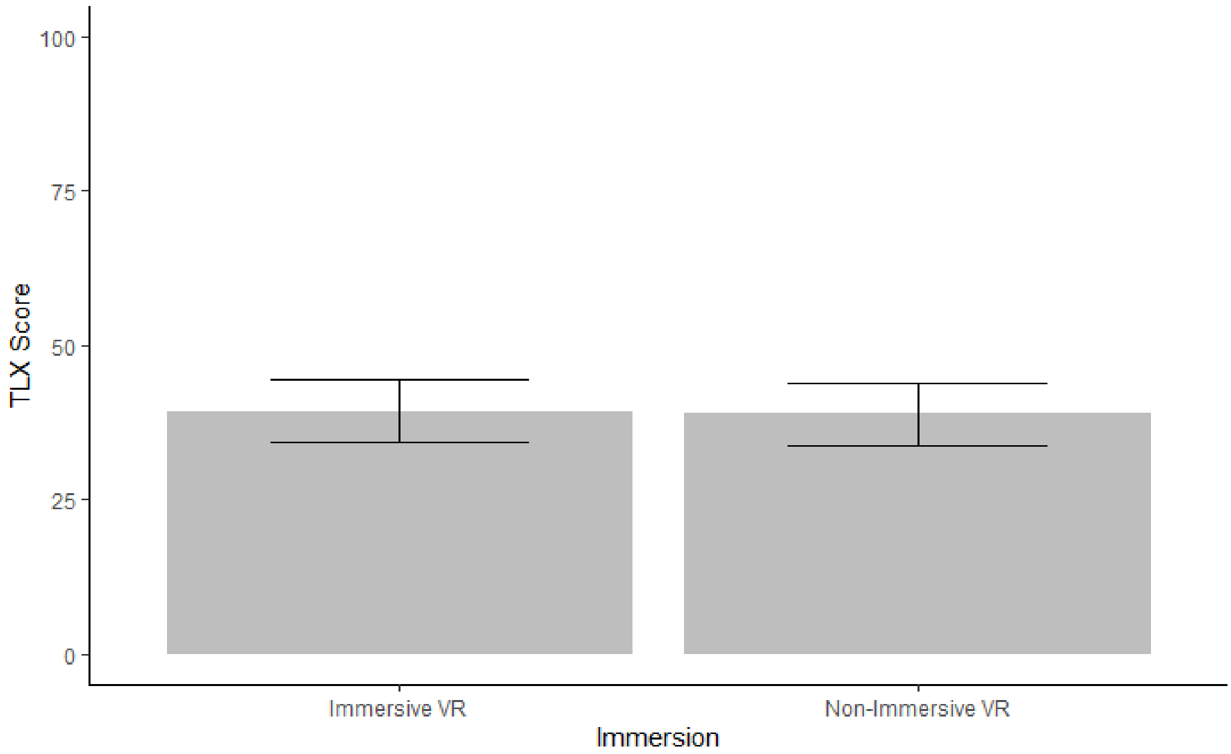


Figure 3.8 TLX Score as a Function of Immersion Condition. Error bars indicate 95% confidence interval.

Hypothesis 5. To test H5, another Tukey’s HSD was conducted to determine the differences in TLX scores between the different interfaces across immersion levels. In nonimmersive virtual reality, no significant differences existed between the visual and visual-auditory interfaces, the visual and visual-haptic interfaces, or the visual-auditory and visual-haptic interfaces (see Table 3.4). However, in immersive virtual reality, the visual-haptic interface outperformed the visual interface and the visual-auditory interface outperformed the visual interface, and there are no significant differences between the visual-auditory and visual-haptic interfaces as can be seen in Figure 3.9. These results support $H5_a$, suggesting that multimodal interfaces are more effective at reducing cognitive load in immersive environments.

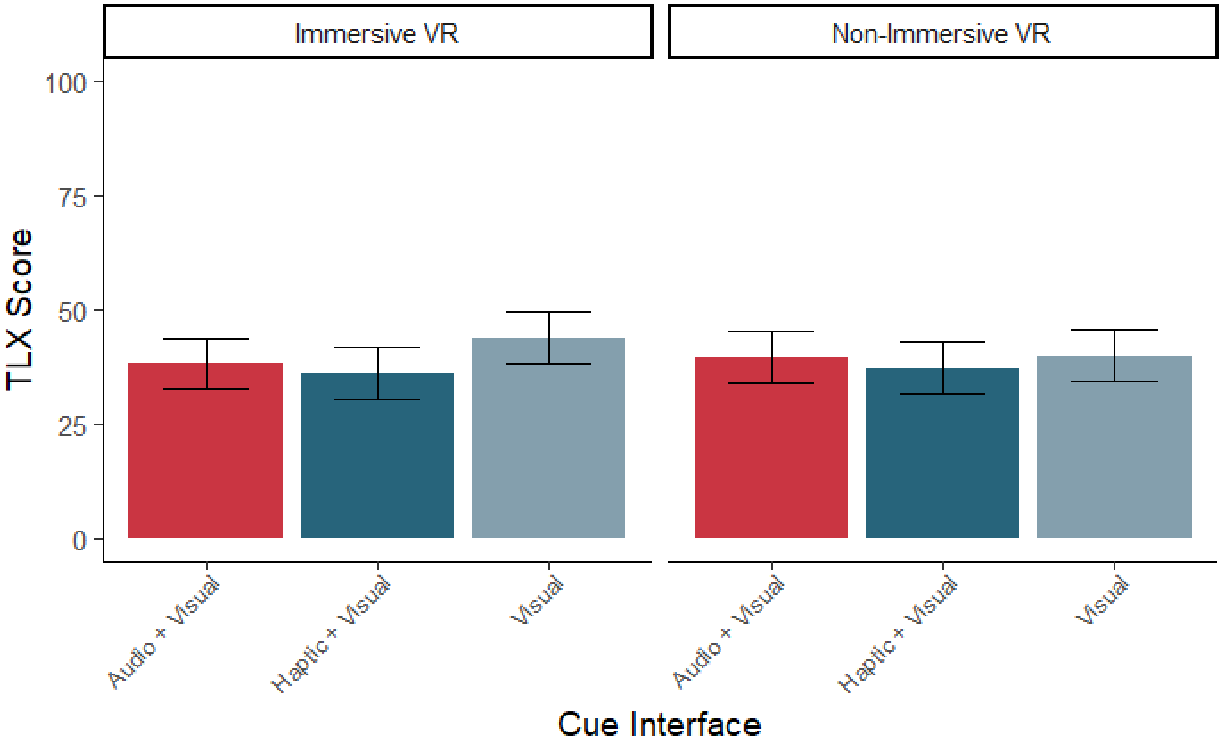


Figure 3.9 TLX Score as a Function of Immersion Condition and Cue Interface Condition.

Table 3.6 Pairwise Comparisons of TLX scores across Cue Interface and Immersion Conditions. Estimates represent the differences in performance across interface types and immersion levels.

Comparison	B	*SE*	*t*	*p*
Non-Immersive VR				
Visual-auditory to Visual Interface	.80	1.83	.44	.90
Visual-haptic to Visual Interface	-1.84	1.83	-1.00	.57
Auditory Visual to Visual-haptic Interface	2.64	1.83	1.44	.32
Immersive VR				
Visual-auditory to Visual Interface	-5.47	1.80	-3.04	.01
Visual-haptic to Visual Interface	-7.36	1.80	-4.09	<.001
Auditory Visual to Visual-haptic Interface	1.89	1.80	1.05	.55

Chapter 4. Discussion

The purpose of this study was to investigate the effects of immersion and multimodal cues on cognitive load, and how the relationship between these constructs varies across difficulty. Within this investigation, I examined the interactions that these constructs have with one another, using both behavioral and self-report measures of cognitive load. This study proposed competing hypotheses. Multiple Resource Theory predicts that the inclusion of an additional task requires additional cognitive ressources to perceive, and so adding additional information in the form of immersion or multimodal cues would increase cognitive load (Wickens, 2008). However, literature on flow theory and redundant cueing suggests that the presence generated from immersive virtual reality and additional cues can alleviate cognitive load (Bjork & Bjork, 2011; Xie *et al.*, 2017). The results of this experiment suggest that IVR and multimodal interfaces decreases cognitive load. This effect is particularly pronounced when the visual cue is accompanied by a haptic cue. The implications of these findings for system design, compatibility with multiple resource theory, and future research directions will be discussed.

4.1 Multimodal Interfaces

The study found a small but significant effect of multimodal interface on cognitive load. The likelihood of shooting the targets in the visual-haptic condition was greater than in the visual interface in an overall comparison. Additionally, TLX scores were significantly lower for the visual-haptic condition than in the visual interface. These findings support the redundant cueing literature by showing a reduction in cognitive load in multimodal interfaces. Literature on redundant cueing suggests that the addition of a redundant cue can assist users in navigating to and focusing on learning material (Xie *et al.*, 2017). This runs contrary to the predictions of MRT, which suggest that the addition of any cues, however complimentary, incur a general cognitive cost that should increase cognitive load with each additional cue (Wickens, 2008). However, the results should be taken cautiously, as the results do not indicate that using multimodal interfaces to decrease cognitive load is practically significant in most situations, as the effect sizes seen in this experiment are quite small.

The actual difference in performance was very small, about 1.5% in performance and 5 points on the TLX. While this difference does appear to exist based upon the statistical significance of the findings, practically, there is little difference between the two types of interfaces. This slight benefit does not validate major investments into redesigning pre-existing interfaces in most cases, but may be a consideration when designing an interface. However, multimodal interfaces can contribute to a positive user experience aside from cognitive load, increasing enjoyment and increasing reaction time (Tang *et al.*, 2022). The results from this study indicate that multimodal interfaces can be utilized for these benefits without fear of introducing additional cognitive load.

No difference was observed in cognitive load across difficulty between unimodal and multimodal interfaces. Performance appeared to deteriorate at a similar rate across all interfaces, indicating that using this manipulation in highly demanding environments is likely to not improve cognitive load to a great degree. The only difference observed was a very small effect between the visual-haptic and visual-auditory interfaces (see Table 3.2). While the visual-haptic interfaces did improve cognitive load over the visual interface, no significant effect was seen for the visual-auditory interface. By the principles of multiple resource theory and redundant cueing theory, both audio and haptic cues should have the same effect on cognitive load. Because both provide an additional nonvisual modality, concurrent processing with a visual task with either cue should not experience as much interference as a visual task would (Wickens, 2008; Xie *et* *al.*, 2017). Thus, it was expected that the visual-auditory and visual-haptic interfaces would both affect cognitive load in a similar way.

4.2 Immersion

Immersion was found to significantly increase performance by about 6% compared to the nonimmersive group and the difference between the two conditions grew as difficulty increased. These results appear to support the hypotheses advanced by flow theory, that people are able to better achieve a flow state under increased cognitive load. However, it is difficult to know whether immersion (and consequently presence) itself decreased cognitive load. Literature on immersive virtual reality has found that more immersive environments are more cognitively demanding than nonimmersive virtual reality due to the increase of information (Frederkson *et al.*, 2020). If it is true that immersive virtual reality increases cognitive load and that the flow state did increase performance in this environment due to an increase in presence, it follows that the flow state must have increased performance a substantial amount in order to offset the increased cognitive load and increase performance significantly. However, there is reason to be skeptical of this conclusion. While the performance difference may indicate that cognitive load was lower in the immersive condition, no significant differences were found in TLX scores between the two conditions. This is likely attributable to the very small effect size of this interaction.

4.3 Limitations

Cognitive load was measured through performance in this experiment. While performance has been shown to be a valid measure of cognitive load, it is unknown whether the relationship between performance and cognitive load was consistent across all modalities (Haji *et al.*, 2015). For example, one study found that the combination of haptic and visual information increased reaction time compared to visual interfaces while not reducing cognitive load (Riggs *et al.*, 2017). It was expected that the visual-auditory and visual-haptic interfaces would both affect cognitive load in a similar way, but the results from the experiment failed to show this. There are a few possible reasons for this discrepancy. First, the visual, audio, and haptic cues were not completely analogous to one another. While all cues were designed to be similar in duration and intensity, the cues were qualitatively different, with the audio cue attempting to provide a semi-realistic gun-readying sound. While this design decision was made because real virtual reality games often have realistic sound effects, the sound effect is not analogous to either the visual or haptic cue. The theories that informed in the hypotheses of this study predicted no significant differences between auditory and haptic cues, and this inequality between cue modalities may have contributed to this difference. Future studies should take into consideration this issue and design an auditory cue that is more analogous to the visual and haptic cues to rule out this possible confound.

4.4 Future Directions

Given that the primary task for this experiment was strongly related to reaction time, further studies should be conducted to determine if the results generalize to a more diverse set of primary tasks. Furthermore, the secondary task (the cue) in this game was quite simple and required little processing. If a task requires more complex processing in order to functionally use, the concurrent processing may lead to more disruptive resource sharing (Wickens, 2008). Therefore, it appears as though the cost of ignoring cues that are more difficult to process may be less than the cost of processing those cues alongside a more optimal cue. Interfaces with primarily unimodal cues may benefit by adding in an additional cue of a different modality for users to use when it is less optimal to use the primary cue.

The interaction between the multimodal interfaces and immersion revealed that the majority of the differences in cognitive load between cue interfaces existed in the immersive condition. No significant differences were seen between interfaces in the nonimmersive condition. If immersion did in fact increase cognitive load, these results strongly support the hypothesis that the effectiveness of multimodal interfaces increased due to the increased visual demand of virtual reality providing more benefit to interfaces that used a non-visual modality. However, as previously stated, it is unknown whether or not immersive virtual reality actually increased cognitive load. It is difficult to draw conclusions about the effectiveness of multimodal interfaces across all immersive environments without knowing the impact of immersiveness on cognitive load. Future research should investigate the impact of individual immersive factors, like framerate, resolution, and field of view in addition to headset virtual reality to understand how these factors differentially impact cognitive load across different types of multimodal interfaces (Cummings & Bailenson, 2015).

Chapter 5. Conclusion

This study delved into the impact of multimodal interfaces and immersive virtual environments on cognitive load during a first-person target shooting game. The results indicated that both immersive environments and multimodal interfaces effectively reduced cognitive load, with multimodal interfaces showing particular efficacy in immersive settings. Despite a small but significant improvement in performance with multimodal interfaces, caution is advised in interpreting the practical significance of this finding. The study revealed that immersion, as provided by virtual reality, significantly increased performance, but the relationship between immersion and cognitive load remains uncertain. While the study's limitations suggest the need for further exploration, these findings also suggest that incorporating additional modalities in interfaces, especially in highly immersive environments, may be beneficial for enhancing performance and reducing cognitive load. However, the results that were found with this study do suggest that interfaces created with multimodal interfaces do slightly decrease cognitive load. When designing an interface for situations with an intrinsically high cognitive load that require a high amount of processing with a particular modality, the small decrease in cognitive load may be valuable. For instance, in military training simulations, multimodal interfaces should be preferred to unimodal interfaces if possible. Still, it is important consider how little the effect is; the inclusion of an additional cue should supersede tan existing, well-designed unimodal interface. This research contributes valuable insights for system design and underscores the complexity of the relationship between immersion, multimodal interfaces, and cognitive load.

References

Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education, 166*, 104154. doi:10.1016/j.compedu.2021.104154

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.6-5. Retrieved from https://CRAN.R-project.org/package=Matrix

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society (pp. 56–64*). Worth Publishers.

Chang, C. C., Liang, C., Chou, P. N., & Lin, G. Y. (2017). Is game-based learning better in flow experience and various types of cognitive load than non-game-based learning? Perspective from multimedia and media richness. *Computers in Human Behavior, 71*, 218-227.

Chiossi, F., Welsch, R., Villa, S., Chuang, L., & Mayer, S. (2022). Virtual reality adaptation using electrodermal activity to support the user experience. *Big Data and Cognitive Computing, 6*(2), 55.

Csikszentmihalyi, M. (1975). Beyond Boredom and Anxiety. Washington: Jossey-Bass Publishers.

Csikszentmihalyi, M. (1997). Flow and the psychology of discovery and invention. *HarperPerennial, New York*, *39*, 1-16.

Cummings, J. J. & Bailenson, J. N. (2015). How immersive is enough? A meta-analysis of the effect of immersive technology on User Presence. Media Psychology, 19(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Flemisch, FO, Bengler, K, Bubb, H, Winner, H, & Bruder, R (2014) Towards cooperative guidance and control of highly automated vehicles: H-Mode and Conduct-by-Wire. *Ergonomics, 57*(3), 343-360

Frederiksen, Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., & Andersen, S. A. W. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. Surgical Endoscopy, 34(3), 1244–1252. https://doi.org/10.1007/s00464-019-06887-8

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression, 3rd Edition*. Thousand Oaks, CA. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html

Gallegos-Nieto, E., Medellín-Castillo, H. I., González-Badillo, G., Lim, T., & Ritchie, J. (2016). The analysis and evaluation of the influence of haptic-enabled virtual assembly training on Real Assembly performance. The International Journal of Advanced Manufacturing Technology, 89(1–4), 581–598. https://doi.org/10.1007/s00170-016-9120-4

Ghani, Jawaid A., Roberta Supnick, and Pamela Rooney. "The experience of flow in computer-mediated and in face-to-face groups." (1991).

Granados, J., Canare, D., & Vangsness, L. (in press). Level-Up! Comparing accessiblity features on gameplay performance. Journal of Accessibility and Design for All

Hashemian, A. M., Adhikari, A., Aguilar, I. A., Kruijff, E., Heyde, M. V., & Riecke, B. E. (2023). Leaning-Based Interfaces Improve Simultaneous Locomotion and Object Interaction in VR Compared to the Handheld Controller. IEEE transactions on visualization and computer graphics, PP, 10.1109/TVCG.2023.3275111. Advance online publication. https://doi.org/10.1109/TVCG.2023.3275111

Infinite.cz. HMD Tester Virtual Reality Headset Database Utility. Infinite.cz. https://www.infinite.cz/projects/HMD-tester-virtual-reality-headset-database-utility

Kiili, K. (2004). Digital game-based learning: Towards an experiential gaming model. Internet and Higher Education, 7(4), 291–302. https://doi.org/10.1016/j.iheduc.2004.12.001

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning [Editorial]. Learning and Instruction, 12(1), 1–10. https://doi.org/10.1016/S0959-4752(01)00014-7

Lenth, R. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.10.0. Retrieved from https://CRAN.R-project.org/package=emmeans

Lu, S. A., Wickens, C. D., Sarter, N. B., & Sebok, A. (2011). Informing the Design of Multimodal Displays: A Meta-Analysis of Empirical Studies Comparing Auditory and Tactile Interruptions. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55(1), 1170-1174. https://doi.org/10.1177/1071181311551244

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60,* 225–236.

Marucci, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E. F., ... & Aricò, P. (2021). The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific Reports, 11*(1), 4831.

Pearce, J. M., Ainley, M., & Howard, S. (2005). The ebb and flow of online learning. *Computers in human behavior, 21*(5), 745-771.

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Robertson, G. G., Card, S. K., & Mackinlay, J. D. (1993). Three views of virtual reality: non immersive virtual reality. *Computer, 26*(2), 81.

Schrader, C., & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. Computers in Human Behavior, 28(2), 648–658.

Slater, & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. Presence : Teleoperators and Virtual Environment, 6(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Strasburger, H. (2020). Seven Myths on Crowding and Peripheral Vision. *I-Perception, 11*(3). https://doi.org/10.1177/2041669520913052

Unity. (2023). *Unity* [Game Building Software]. https://unity.com/

VR Games, apps, & videos. VIVEPORT. (n.d.). https://www.viveport.com/

Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in human behavior, 9*(4), 411-426.

Weaver, B. W., & DeLucia, P. R. (2022). A Systematic Review and Meta-Analysis of Takeover Performance During Conditionally Automated Driving. Human Factors, 64(7), 1227–1260. https://doi.org/10.1177/0018720820976476

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, *50*(3), 449-455.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. Retrieved from https://CRAN.R-project.org/package=dplyr

Xie, H., Wang, F., Hao, Y., Chen, J., An, J., Wang, Y., & Liu, H. (2017). The more total cognitive load is reduced by cues, the better retention and transfer of multimedia learning: A meta-analysis and two meta-regression analyses. *PloS one*, *12*(8), e0183884. https://doi.org/10.1371/journal.pone.0183884