NAVIGATION PERFORMANCE IN COMPLEX MULTI-STORY ENVIRONMENTS USING AUDIO-FIRST MIXED REALITY

by
BİLGEHAN ÇAĞILTAY

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of Master of Science

Sabancı University
June 2025

NAVIGATION PERFORMANCE IN COMPLEX MULTI-STORY ENVIRONMENTS USING AUDIO-FIRST MIXED REALITY

Approved by:

Prof. Dr. Selim Balcısoy ....
(Thesis Supervisor)

Assoc. Prof. Dr. Mehmet Göktürk ....

Assoc. Prof. Dr. Selçuk Artut ....

Date of Approval: June 19, 2025

ABSTRACT

NAVIGATION PERFORMANCE IN COMPLEX MULTI-STORY ENVIRONMENTS USING AUDIO-FIRST MIXED REALITY

BİLGEHAN ÇAĞILTAY

COMPUTER SCIENCE & ENGINEERING MSc. THESIS, JUNE 2025

Thesis Supervisor: Prof. Dr. Selim Balcısoy

Keywords: MR, situated analytics, AAR

MR interfaces have traditionally prioritized visual modalities as the primary conduit for user interaction. While effective in certain contexts, this visual-first approach reveals critical limitations in high-stakes, cognitively demanding environments where sustained visual attention may be impractical or counterproductive. This study challenges the dominance of visual-centric design in MR and advocates for a transformative shift toward audio-centric interaction paradigms. Accordingly, a conceptual framework for an Audio-Based Situated Analytics system is proposed. Such a system, rebalances sensory load by elevating auditory engagement and reserving visual channels for the most urgent, attention-critical tasks. This holds the potential to significantly enhance usability, resilience, and inclusivity in a range of applications. An AR navigation task software is developed by following this proposed framework. Finally, the study experimentally evaluates the effectiveness of 3D spatial navigation in an AR environment using this software. It compares the performance of participants using audio AR navigation to those using visual AR navigation, traditional navigation without aids, and a combination of audio and visual AR. The results of this study show that audio AR has the capability to maintain the environmental awareness of a user similar to that of a person who does not use any AR assistance. The results also show that audio AR is capable of having similar performance to visual AR in the navigation of complex, 3D environments. The results show similar findings for audio AR in 3D environments to those of previous literature, which were mainly focused on 2D environments. As a conclusion, this study provides several insights on the benefits of audio AR in the context of navigation of 3D environments. The results also identify several areas for future research.

ÖZET

KARMAŞIK ÇOK KATLI ORTAMLARDA SES ÖNCELİKLİ KARMA GERÇEKLİK KULLANILARAK NAVİGASYON PERFORMANSI

BİLGEHAN ÇAĞILTAY

BİLGİSAYAR BİLİMİ VE MÜHENDİSLİĞİ YÜKSEK LİSANS TEZİ, HAZİRAN 2025

Tez Danışmanı: Prof. Dr. Selim Balcısoy

Anahtar Kelimeler: KG, konumsal analiz, SAG

KG arayüzleri geleneksel olarak kullanıcı etkileşimi için görsel modalitelere öncelik vermiştir. Çoğu bağlamda etkili olsa da, bu yaklaşım bazı durumlarda kritik sınırlamalara yol açar. Sürekli görsel dikkat gerektiren, yüksek riskli ve bilişsel olarak zorlayıcı ortamlar buna örnektir. Bu çalışma KG'de görsel merkezli tasarıma alternatif olarak ses merkezli bir etkileşim modeli önermektedir. Böyle bir sistem, işitsel katılımı artırarak ve görsel kanalları en acil görevler için ayırarak duyusal yükü dengeler. Bu yeniden kavramsallaştırma, birçok uygulamada kullanılabilirliği, dayanıklılığı ve kapsayıcılığı artırma potansiyeline sahiptir. Örneğin görsel alanı boğmayan, durumsal farkındalık gerektiren, zamana duyarlı ilk müdahale operasyonları ve sürükleyici kültürel miras deneyimleri bu ortamlardandır. Ayrıca, önerilen çerçeveye uygun bir AG navigasyon görev yazılımı geliştirilmiştir. Bu yazılımla 3B mekansal sesli navigasyonun etkinliği deneysel olarak değerlendirilmiştir. Sesli AG navigasyonu kullanan katılımcıların performansı, görsel AG navigasyonu, geleneksel navigasyonu ve sesli ve görsel AG'nin kombinasyonunu kullananlarla karşılaştırılmıştır. Sonuçlar sesli AG'nin geleneksel navigasyona benzer çevresel farkındalık sağladığını göstermektedir. Ayrıca, karmaşık 3B ortamlarda görsel AG ile benzer performans sunabildiği görülmüştür. Buna ek olarak, genellikle 2B ortamlara odaklanan önceki literatürdeki bulgulara paralel olarak, 3B ortamlarda da sesli AG'de benzer sonuçlar elde edilmiştir. Sonuç olarak bu çalışma, 3B ortamlarda navigasyon bağlamında sesli AG'nin faydaları hakkında çeşitli içgörüler sunmaktadır. Sonuçlar ayrıca bu alanda ihtiyaç duyulan bir çok araştırma için de yönlendirmeler içermektedir.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support and guidance of my mentors, friends, and family. I am very grateful for the guidance I have received from my supervisor, Prof. Dr. Selim Balcısoy. His expertise, support, and guidance along my research journey has been monumental. Without his input, I would have been lost among the literature. I have learned how to keep myself on track, and have become a competent scientist thanks to his guidance.

I would also like to thank my jury members, Assoc. Prof. Dr. Mehmet Göktürk and Assoc. Prof. Dr. Selçuk Artut, for their time and their invaluable comments on my thesis.

I'm thankful for the support I have received from my friends who provided their emotional and technical support throughout my journey. I would especially like to thank İ. Efsane Algın for her expertise and input regarding the measurement of attention.

Finally, I want to thank my loving family for their support and encouragement throughout my Master's program. I would not have been able to complete my education successfully without their willingness to be there when I needed them.

Ad Astra, Per Aspera

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATONS

2B	İki Boyutlu	v
2D	Two Dimensional	iv, 9, 55
3B	Üç Boyutlu	v
3D	Three Dimensional	iv, 6, 8, 9, 11, 25, 37, 50, 55
6DOF	6 Degrees of Freedom	x, 22, 23, 32, 34
AAR	Audio Augmented Reality	iv, 3, 4, 11, 12, 14
AG	Arttırılmış Gerçeklik	v
ANOVA	Analysis of Variance	39, 40, 43
AR	Augmented Reality	iv, 4, 5, 6, 7, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 55, 56
FASS	Faculty of Arts and Social Sciences	xi, 24, 26, 27
FENS	Faculty of Engineering and Natural Sciences	15
FOV	Field of View	1, 3
FR	First Responder	1, 2, 9, 10
HMD	Head Mounted Display	6, 7, 12
HRTF	Head Relative Transform Function	4, 9, 11, 58
KG	Karıştırılmış Gerçeklik	v
MR	Mixed Reality	iv, 1, 2, 3, 4, 7, 9, 10, 11, 12, 13
MRTK	Mixed Reality ToolKit	20
NASA	National Aeronautics and Space Administration	19, 28, 43, 52

PC Personal Computer 20, 22, 23, 36

SAG Sesli Arttırılmış Gerçeklik v

UNIX Uniplexed Information Computing System 33, 34

VR Virtual Reality 6, 7, 8, 9, 12, 13, 15, 16, 46

1. INTRODUCTION

MR interfaces have a unique set of applications with many innovative potentials. The strides made in display technologies in the past twenty years have drastically expanded the boundaries of what is achievable. This has created a very valuable technological innovation, the Hololens 2, which is still in popular use after six years of its release. MR interfaces are vision and hand interaction focused technologies. As a result, historically MR research has mainly focused on visual interfaces to facilitate information delivery and user interaction. This has allowed the research in the field to apply digital information overlays easily and reliably for complex, technical tasks. On top of what was visible to normal human perception, it was possible to now display anything desired; ranging from displaying infrared colors to showing the internal structures of complex machines, like a car engine. It further allowed the feasibility of remote working with experts in hard to reach locations or training with expensive equipment without the risk of any financial losses caused by damage during training. Finally, it became much more feasible to augment the perception of people in high-stress environments, where receiving situation critical information can be a matter of life and death. This was all done by giving users a digital information layer in their visual fields.

However, visual MR systems come with their own trade-offs. Hand tracking, eye tracking, and voice commands have emerged as primary methods for interacting with these interfaces [28]. While hand-tracking systems are effective in typical everyday contexts, they face considerable limitations in extreme environments, such as those encountered by FR teams who may be required to wear thick protective gloves for fire safety, or for medical emergencies where hand interaction is often difficult [60]. These limitations are further exacerbated by factors such as the limited FOV and screen resolution, both of which play critical roles in determining the effectiveness of MR systems in real-world contexts. For example, MR headsets like the Hololens 2, with its 52° horizontal FOV, can be restrictive, limiting what can be shown at a given time. Additionally, the challenge of balancing attention between virtual and real-world information can distract users, especially in high-pressure scenarios, reducing the effectiveness of MR in these settings [15; 4].

Furthermore, the inherent limitations of projected MR displays, especially in brightly lit environments, further hinder their ability to deliver information effectively. The inability to see displays clearly in direct sunlight or harsh lighting conditions can impede the ability to complete time-sensitive tasks, increasing user frustration and cognitive load [30]. In addition, due to the projected nature of the display, these devices need to be fairly large in their form factor, limiting their wear-ability and applicability. These visibility and usability issues contribute to diminished immersion, greater mental fatigue, and ultimately, a decline in operational efficiency of MR in critical applications.

It can be seen that the continued exclusive reliance on visual modalities may be constraining the broader potential of MR. This predominantly visual approach not only risks overlooking alternative interaction methods, but also limits usability, accessibility, and cognitive efficiency, undermining MR's capacity to deliver in a wider array of real-world applications. Moreover, the advancing computational power of MR devices, now available in increasingly compact and powerful form factors, presents unique opportunities for extending prior research. This burgeoning computational capacity allows for more sophisticated MR interfaces that can overcome many of the usability challenges previously faced by users relying on vision-heavy interfaces. By harnessing this enhanced computational power, it becomes possible to refine and extend the functionalities of MR systems, addressing longstanding limitations.

To mitigate these challenges, we propose a shift away from a visual-dominant interface, advocating for a more integrated use of sensory modalities. This study first proposes a shift in MR interface design, advocating for the prioritization of audio-centric interfaces. By challenging the entrenched visual-first paradigm, this approach opens the door to more versatile, contextually relevant, and immersive MR experiences.

FR, such as police officers, paramedics, and firefighters, often find themselves in high-stakes situations that demand both hands-free engagement, as well as full cognitive and visual attention; all while needing to access crucial, real-time situational data. In these critical scenarios, auditory interfaces may provide a more effective and efficient means of conveying essential information without detracting from the task at hand. This interface can display friend or foe identifier information for users in areas with adversaries, or convey navigational information in complex spaces where users cannot effectively navigate, due to unfamiliarity with the environment or reduced visibility in cases of fire.

This also has the potential to elevate the daily experiences of civilians. Users in unfamiliar environments that may need navigational assistance can use verbal or non-verbal audio guidance for hands free navigation in areas like large airports or complex building interiors. Furthermore, this can be used to aid civilians who may have difficulties in visual navigation.

Besides these applications, cultural heritage sites stand to benefit significantly from this approach, enhancing tourist experiences through the use of AAR [54], offering a deeper and more interactive engagement with historical artifacts and sites. This research highlights the audio modality as an under-explored but valuable alternative. Unlike vision, which is constrained by line of sight, hearing offers a 360° FOV, providing a more holistic and flexible means of interaction. This auditory advantage has the potential to alleviate many of the cognitive and visibility-related issues associated with visual MR systems.

The spatial nature of hearing enables users to engage with objects throughout their entire environment, not limited to what is directly in front of them. In high-stress situations, where critical information may originate from behind the user, relying on visual cues can be both distracting and disruptive. By contrast, auditory cues allow for seamless, non-intrusive interaction, empowering users to remain fully engaged with the environment while receiving vital information. In addition, due to the reduced form factor, from bulky glasses to a simple pair of earbuds, it is easier to move around with an audio-only interface.

Such an audio system can be implemented through utilizing methods researched within the field of sonification. Sonification is the use of non-speech audio to convey information or perceptualize data. As such, an audio system can be implemented through utilizing methods researched within the field of sonification. Unlike general auditory cues or alerts, sonification involves mapping data to structured auditory signals in a way that users can interpret meaningfully. For example, audio properties such as pitch, tempo, or timbre can be varied to represent different types of information, like distance to a target, speed, temperature, or abstract analytics like risk levels or performance metrics. Similarly, audio icons such as a ping, a bell ringing, or a beeping sound can also be used to implement cues of events with real-life analogs [52; 41; 6]. By utilizing sonification within MR, audio can be conveyed with its spatial properties, as well as any information that may be encoded within the audio itself.

Sonification and MR share the core goal of augmenting human perception with digital information, but they diverge in modality: MR is a technology that specifically utilizes or augments the spatial environment of its user, while sonification does not require such a focus, since it is interested in how audio can be used to convey complex information and does not necessarily require spatial augmentation. As a result, sonification and MR can complement each other to form a richer, multimodal MR experience.

Hence, further research is essential to compare the efficacy of audio-only MR interaction with traditional methods such as map inspection or visual-only MR interaction. Additionally, there is a need to explore how users interact with and learn to navigate complex, multi-floor environments using auditory cues alone. This includes investigating how audio-only guidance impacts user orientation, decision-making, and overall navigation efficiency in spaces with multiple floors or intricate layouts, which could provide deeper insights into the feasibility and potential advantages of audio-centric approaches in MR applications.

Leveraging of HRTF and AAR technologies offer significant promise in enabling users to perform spatial orientation and navigation tasks using auditory direction alone. By creating spatial audio cues that indicate the location of critical information, users can efficiently navigate complex environments without the need for visual input [46]. The integration of auditory information not only has the potential to enhance comprehension but it also allows for the reduction of visual clutter, leaving space for the most important elements of the visual interface [43; 37]. As a result, users can engage more effectively and intuitively with the MR environment, leading to improved performances.

Furthermore, audio displays require far fewer resources, only two high-quality speakers, compared to visual displays that demand a broad field of view and high resolution to convey equivalent levels of detail. This not only makes auditory interfaces more cognitively efficient, but also more practical and accessible, especially in resource-constrained or high-pressure contexts.

The primary goal of this study is to develop an innovative audio-based situated analytics system. This will be done by integrating situated analytics and AAR to let users interact with environmental data in a hands-free, vision-free manner. By harnessing the power of this technology, users will gain access to real-time critical information, enhancing their decision-making capabilities and situational awareness without the need for visual distractions.

The integration of AAR and location analytics represents a cutting-edge approach to AR. This study aims to break free from the traditional reliance on visual modalities by incorporating aural cues, enabling users to bypass the limitations of vision and interact with their surroundings more efficiently. While current AR systems primarily focus on visual interfaces, this approach utilizes the unique advantages of auditory feedback, offering a more flexible, immersive, and efficient user experience, particularly in environments where visual attention is constrained or overloaded. To this end, this study will compare the performances of participants with audio-only, visual-only, audio-visual, and no AR systems.

Accordingly, this study aims to answer three main research questions, and two sub-research questions for each.

RQ1. How does AR impact navigational performance, evaluated by time and distance?

RQ1.1. Does audio AR improve navigational performance?

RQ1.2. Does audio AR decrease navigational performance?

RQ2. How does AR impact environmental awareness in navigation?

RQ2.1. Does audio AR improve environmental awareness in navigation?

RQ2.2. Does audio AR decrease environmental awareness in navigation?

RQ3. How does AR impact task load of tasks in navigation?

RQ3.1. Does audio AR improve task load of tasks in navigation?

RQ3.2. Does audio AR worsen task load of tasks in navigation?

2. BACKGROUND OF THE STUDY

VR and AR have emerged as transformative technologies that are reshaping human interaction with digital environments. While they are often discussed together due to their shared goal of altering perception, their historical roots, technological developments, and research trajectories reveal distinct paths.

2.1 History of VR and AR

VR refers to a fully computer-generated simulation in which users can interact with a three-dimensional environment using specialized hardware such as HMD, gloves, and motion trackers. The concept of simulating reality dates back several decades, with early attempts aimed more at mechanical illusions than digital immersion.

The conceptual groundwork for VR began as early as the 1930s. In [57], the story described a pair of goggles that could provide the wearer with a fully immersive experience, complete with visuals, sound, smell, and touch. This fictional notion foreshadowed the eventual goals of VR.

In technological terms, one of the earliest precursors to VR was [23], a multi-sensory simulator that offered stereo sound, stereoscopic 3D visuals, vibrations, and even scents. Though it was mechanical rather than digital, it represented the first attempt to create a holistic immersive experience. A recreation of this device has also been made for preservation purposes [1].

The development of HMD's was a significant leap forward. In 1968, Ivan Sutherland, often referred to as the "father of computer graphics," created the first HMD system known as the Sword of Damocles [58]. It was connected to a computer and allowed users to see simple wireframe environments overlaid in their vision. Despite its crude graphics and large mechanical support, it was the first true digital VR system. This device is also argued to be the first recorded use of the term "HMD".

Throughout the 1980s and 1990s, the concept of VR gained further traction with the work of Jaron Lanier, who founded VPL Research, one of the first companies to sell VR products like the DataGlove and EyePhone HMD [55]. The term "virtual reality" was popularized during this era and VR began to see use in military training, flight simulation, and academic research.

Early academic research on VR focused on user interface design, navigation in virtual environments, and presence, which describes the feeling of "being there." A foundational paper by Slater [48] emphasized the psychological and perceptual dimensions of VR, attempting to quantify what makes users feel immersed.

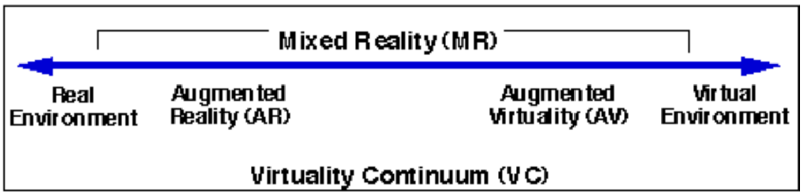
In contrast to VR, AR involves overlaying digital information onto the real world. Rather than replacing reality, AR enhances it by superimposing graphics, sounds, or other sensory enhancements. This fundamental difference shapes both its applications and its technological requirements.

Similarly to VR, the conceptual roots of AR date back to the mid-20th century. The term "augmented reality" was coined by Boeing researcher Tom Caudell to describe a digital display system that guided workers through complex manufacturing tasks [11]. However, AR had already started being explored in earlier works, such as Sutherland's 1968 HMD, which could be considered the first AR system because it overlaid virtual graphics onto the physical world.

One of the first widely recognized AR systems was developed by Ronald Azuma [5], who defined AR as systems that combine real and virtual objects in a real environment, are interactive in real time, and register/align the virtual and real objects with each other. Azuma's paper marked a foundational point in AR research, categorizing early AR systems and identifying the key technical challenges of registration accuracy, latency, and display quality.

A foundational framework for understanding the relationship between reality and virtuality is the virtuality continuum [39]. This continuum, shown in Figure 2.1, conceptualizes a spectrum between the real environment at one end and a fully virtual environment at the other. Between these poles lie various MR experiences, including AR and Augmented Virtuality, where real-world elements are inserted into a primarily virtual space.

Figure 2.1 Virtuality continuum [39]



This continuum was developed to clarify the distinctions between different types of mediated reality experiences, particularly as early systems began blending digital and real-world content. It helped researchers and developers identify where their systems fell in terms of immersion and realism, thus informing design decisions about user interaction, display technologies, and tracking mechanisms.

According to this model, AR lies closer to the real-world end of the spectrum. This is because the real environment remains dominant, but virtual elements are added to it. Augmented Virtuality lies closer to the VR end. Here, a primarily virtual environment incorporates real-world content (e.g., video feeds or real-time sensor data).

The virtuality continuum continues to be a key theoretical framework in AR/VR research and design, particularly in the fields of education, industrial training, and entertainment, where different degrees of immersion and interactivity are required.

AR technologies evolved rapidly in the early 2000's with the advent of mobile computing and computer vision. Fiducial marker tracking systems like ARToolKit [33] allowed researchers and hobbyists to experiment with overlaying 3D models onto printed markers through a webcam feed, laying the groundwork for mobile AR applications. Eventually, platforms like Microsoft HoloLens [42] and mobile apps like PokÃĺmon GO brought AR into mainstream consumer use.

Early AR research emphasized tracking and registration, ensuring that virtual content accurately aligns with the physical environment. Works such as [29] tackled these technical hurdles and proposed systems for real-time interactive AR experiences.

2.2 Audio in AR and VR

While visual immersion is typically emphasized in AR and VR systems, sound is an equally vital component in creating convincing and engaging experiences. Sound in AR and VR can be both narrative and functional. They can be used to convey information, guide attention, and deepen emotional engagement.

In VR, spatial audio is particularly important. 3D audio techniques are used to crudely simulate how sound behaves in a physical environment, giving users cues about distance, direction, and the characteristics of virtual spaces. Techniques such as binaural audio, HRTF's, and ambisonics enable sound to dynamically change based on the user's head movements, enhancing realism and presence.

Research into auditory perception in virtual environments began to gain prominence in the 1990s. One notable work is [7], which investigated how spatial audio could enhance realism and user experience. Other studies, such as [8], explored psychoacoustic principles to understand how humans localize sound sources. These provided insights that were essential in designing AR/VR audio systems.

Sound is rarely used in isolation in AR and VR. Its power is maximized when integrated with visual and interactive elements. However, standalone audio AR is an emerging field. Applications like Microsoft's Soundscape [2] provide navigational aid to visually impaired users through spatialized audio cues, demonstrating the potential of sound-driven AR interfaces.

The growing interest in immersive audio is evidenced by the development of standards and tools such as Facebook's (Meta's) Spatial Workstation [21], Google's Resonance Audio[36], and Apple's spatial audio capabilities in ARKit. Academic conferences like the Audio Mostly series and IEEE's VR and 3D User Interfaces often feature dedicated tracks on spatial sound, reflecting a robust and deepening research community.

While extensive research has been conducted on sound localization in stationary contexts and the utility of auditory cues for navigating 2D environments such as street navigation [13; 25; 19; 35; 18] and single-floor navigation in buildings [46; 56; 61], few studies have explored the challenges and potential of 3D terrain navigation with realistic sound simulation [44]. Despite the growing interest in auditory aids for orientation and navigation, there remains a significant gap in research that addresses the complexities involved when users are navigating more dynamic, 3D spaces. The intricacies of simulating 3D sound within these environments could potentially offer new avenues to improve user experience and navigational efficiency, but it has not yet received the level of attention that other aspects of auditory navigation have.

Previous studies on FR using MR interfaces have highlighted a critical concern:

MR systems have the potential to obscure the real environment, which can pose significant risks to users in high pressure and time-sensitive situations [4]. Moreover, the effectiveness of MR displays can be severely compromised in bright or outdoor conditions, as the displays may become difficult to see, hindering the ability of FR's to access vital information when they need it most. Furthermore, there is a level of difficulty in integrating visual MR technologies into existing gear and the daily lives of people, since the display technology itself occupies considerable space.

By transitioning part or all of the information traditionally conveyed through visual channels to the auditory domain, these challenges can be significantly mitigated. When information is delivered through auditory cues, it can reduce mental workload while simultaneously enhancing both audiovisual target acquisition time and perceived situational awareness [4; 37]. This shift not only alleviates the cognitive burden on users but also enables more seamless interactions with the environment, especially in high-pressure situations where visual attention is divided or constrained.

Additionally, audio-only navigation has been shown to help users develop high-quality mental maps more quickly compared to visual-only navigation [13]. These findings are based on preliminary research involving novice users, who demonstrated performance on par with that of visual-only navigation. This suggests that audio cues, even for inexperienced users, can facilitate efficient spatial awareness and navigation, potentially offering an effective alternative to traditional visual guidance in MR environments.

While current evidence is promising, further longitudinal studies are needed to demonstrate that auditory navigation can consistently outperform visual-verbal navigation methods over time [56]. However, the development of audio-assisted or audio-only MR interfaces faces certain limitations, largely due to the relatively limited focus on auditory interaction in past MR research. This under-representation has constrained progress in optimizing audio-based guidance systems, underscoring the need for deeper exploration into auditory interface design and evaluation.

As highlighted by [54], one of the most pressing challenges in audio-based MR interface research is the "lack of formal methodology on how to analyze and interpret user data that is not just qualitative." This methodological gap limits the robustness and generalizability of findings in the field. Moreover, while audio icons offer an efficient means of conveying information, they can easily overwhelm users, particularly when multiple audio sources overlap, making it difficult to accurately localize individual sounds. This perceptual limitation poses a significant usability challenge. However, further research is warranted to explore whether such difficulties can be mitigated through user training, adaptation periods, or the integration of more sophisticated spatial audio techniques.

While audio interfaces offer the potential for comprehensive environmental awareness, the spatial accuracy of auditory cues is not uniformly distributed across all directions. Audio sources located behind the user or in peripheral positions, such as laterally, above, or below, often exhibit reduced localization accuracy and increased perceptual ambiguity. This directional uncertainty is especially pronounced when the listener is stationary, with the greatest difficulty observed in distinguishing between sound sources positioned above versus below the user [54; 16; 40]. Such limitations highlight the need for improved spatial audio rendering techniques and possibly head movement or contextual cues to support more precise sound source localization in MR environments.

Although personalized HRTF speakers can significantly enhance spatial audio accuracy and mitigate localization errors by being specially made for the head shape of its user, they are not yet a viable solution for widespread use in consumer-grade electronics due to cost and complexity. However, these challenges can be partially addressed through user movement and the use of binaural 3D audio with generic HRTF speakers, which allows for improved spatial perception even in the absence of individually tailored HRTF's [49]. Dynamic cues generated through head or body movements enable the auditory system to recalibrate and resolve ambiguities, particularly in distinguishing horizontal sound sources.

Encoding elevation through auditory cues, such as mapping higher frequencies and brighter timbres to higher vertical positions, has been shown to help resolve ambiguities in spatial localization, particularly with respect to elevation [9; 16; 47; 50; 45]. This psychoacoustic strategy leverages the listener's natural associations between pitch and verticality to improve spatial accuracy. However, using these techniques on novice users has the potential to cause confusions during navigational tasks, as they may be confused in whether the sound change is an error that they may have caused. As a result, using these techniques in dynamic environments requires further research. Furthermore, in the context of AAR, more research is necessary to determine the most effective sound types and auditory metaphors for specific tasks, ensuring that audio cues are intuitive, distinguishable, and contextually appropriate.

One study found that the most effective auditory displays combine earcons, structured, non-verbal auditory cues, with 3D spatial audio, enhancing users' spatial awareness and interaction efficiency [54]. However, the study also emphasizes that no single sound type universally outperforms others; rather, the effectiveness of an auditory display largely depends on the distinctiveness of the sound sources and the minimization of overlap, which are critical for avoiding cognitive overload and maintaining clarity in complex auditory environments. Studies have further demonstrated that conversational audio, wherein a narrator directly addresses the user to convey information, significantly increases the cognitive load, potentially detracting from task performance in complex environments [54; 56; 12]. As a result, non-speech auditory options are generally favored when feasible. Users tend to prefer earcons that are intuitively associated with specific tasks, along with symbolic or low-frequency auditory cues that occur sparingly, minimizing distractions while preserving informational clarity.

This highlights the importance of limiting AAR-delivered information to a select set of critical contexts, particularly for novice users who may be more susceptible to cognitive overload. A more restrained and context-sensitive approach to auditory augmentation can enhance clarity and usability, especially in early user experiences. Moreover, placing a stronger emphasis on AAR within consumer-facing applications has the potential to deepen user immersion. For example, in settings such as museums or art galleries, AAR can foster more engaging, informative, and memorable experiences by seamlessly integrating audio narratives or cues into the spatial and thematic layout of exhibits [54].

In such scenarios, users are able to focus more effectively on their surroundings rather than being distracted by persistent visual overlays or textual information. This reduction in visual dependency not only enhances perceptual clarity but also supports a more natural interaction with the physical environment. Furthermore, audio-only interfaces offer a more discreet and lightweight alternative to current MR systems, which often require bulky HMD's or visors. By minimizing the need for obtrusive hardware, these audio-based solutions facilitate seamless integration into everyday settings and encourage broader adoption across diverse use cases [53].

2.3 Taxonomy of AR/VR

Table 2.1 is a taxonomy of the literature in AR and VR technologies. It summarizes the information listed in this chapter and categorizes each study according to its contributions to the literature; VR, AR, and Audio in AR/VR. They are then further divided into their own sub-categories in accordance to what they address.

Table 2.1 Taxonomy table of AR/VR

Category	Sub-Category	Research
VR	History & Tech	[57; 23; 1; 58; 55],
	Presence Research	[48]
AR	Theory & Definition	[11; 5; 39]
	Systems & Applications	[29; 33; 42]
Audio AR/VR	Audio Foundations	[7; 8]
	Toolkits	[2; 21; 36]
	Navigation & Accessibility	[13; 25; 19]
	MR Interfaces	[4; 37]
	Spatial Audio	[54; 16; 40]
	Interaction Design	[12; 53]

3. METHODOLOGY

The aim of this study is to understand the effectiveness of 3D spatial audio navigation in an AR environment. The performance of the participants in Audio AR are compared to the performance of participants who utilize visual navigation in AR, classical navigation with no navigational aids, as well as a combination of audio and visual AR. This comparative user study was conducted between subjects. Qualitative and quantitative data was collected to answer the research questions detailed in section 1. In short, the main research question is on how AR impacts various performance metrics during navigation.

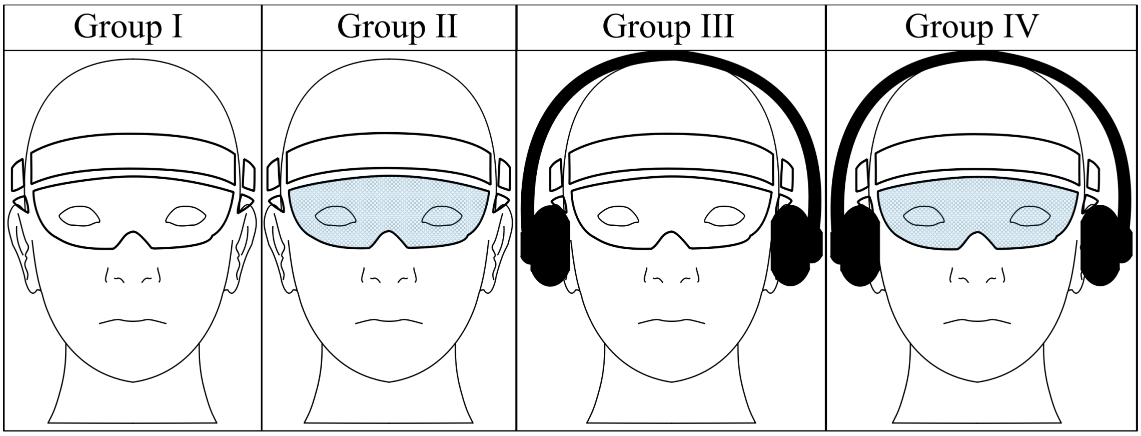
To answer these research questions, an experimental study was conducted in four main groups, as given below:

- Group I: No navigational aids
- Group II: Visual AR
- Group III: Audio AR
- Group IV: Audio + Visual AR

The tasks of each group are explained in more detail in section 3.2. The participants were assigned to one of these groups at random. The randomization was conducted in order of their entrance to the experimental study. That is, the first person was assigned to group I, the second person was assigned to group II, and so on.

To track the movements of all participants, they were all required to wear an AR headset. The participants in group I and II had to only wear the headset, while the participants in group III and IV had to wear the headset, along with a Bluetooth headphone to be able to hear the AAR portion of the task. The distinction between these two groups can be seen in Figure 3.1. Here, the headsets of groups II and IV are shaded to signify that these groups have an active visual AR interface. On the other hand, groups III and IV are wearing headphones to signify that they have an active audio AR interface.

Figure 3.1 Equipment worn by participants in each group. Group II and IV have visual AR, shown by blue highlighted glasses. Group III and IV have audio AR, shown by headphones



3.1 Participants

Each group had a minimum of seven participants, for a total of 28 participants. Each participant was a student of Sabanci University. All participants had a major within the FENS faculty. They were also within the 18-28 age group, with an average age of 22, and no major disabilities in sight, hearing, or attention. A summary of the participants' demographics can be found in Table 3.1. For the experimental study, ethics committee approval was given by Sabanci University.

Ten of the participants had minor disability in sight, as they required glasses. Unfortunately, Group IV had the most amount of participants with any form of glasses use. However, since the usage of the Hololens does not get hindered by the participant wearing glasses, they were informed that they could complete the tasks with their glasses on them. All participants had similar levels of experience with the experiment building. None of them claimed that they would be comfortable in easily finding any given location, besides for specific highly used classrooms. This was expected, which is why these rooms were avoided in this study and the target locations were selected to be unknown spots for normal students.

Group IV had the highest prior experience with AR and VR technologies. As explained in section 3.2, in anticipation of this, all participants were asked to complete an extra task with a dummy location, where they were informed that they could test the system and ask any question they may have. This ensured that all participants were as comfortable with the system as possible.

Table 3.1 Participant demographic summary

Category	Group I	Group II	Group III	Group IV
Total participants	7	7	7	7
Female count	3	4	2	2
Average age	23	21	22	22
Glasses count	1	1	3	5
Average VR/AR experience (5-point scale)	2.1	2.1	1.6	3.2
Average environment familiarity (5-point scale)	3.1	3	3	3.2
Average navigational skill perception	3.7	3.4	3	3
Average gaming rate	3.5	2.5	4	3.8
Navigational sport count	2	1	2	1

3.2 Task Descriptions

Each group is required to complete four navigation tasks, where they reach a specific target location, as well as a dummy location. Each task starts with the participants being asked to navigate to the dummy location. The purpose of the dummy location step is to ensure that the participant fully understands the experimental study before starting the actual experiment, where the full data collection takes place. No assessments are made during this dummy navigation step.

Following this, the participants are assigned to the target locations in a random order. For each target, there is an inconspicuous physical item on the path of the participant, placed in the real world, and used to measure their environmental awareness. While the participant navigates to their target location, they are prompted by the researchers on whether they are aware of these objects in their environment. If the participant has passed by an existing object, the question will be asked within the next three seconds.

It is possible to get participants who are not aware of the object but state that they have seen it, or participants who may answer positively due to suggestion-induced false memory. To control for these cases, participants can also be asked whether they have seen an object that is not actually in their surrounding environment.

The combination of these two methods allows researchers to reliably tell whether a participant is really aware of their surroundings. In total, a participant is prompted with one existing and three non-existing objects per navigation task. This method ensures that researchers can assess a participant's awareness of their surroundings without complications arising from their ability to recall noticing the items. The list of items that were asked to the participants can be found in Appendix B. Additionally, the distribution of these items on the path of the participants, as well as the ideal path of each navigation task can be found in appendix D.

When participants complete their navigation to all four target locations, they are given a set of questionnaires. These questionnaires are explained in detail in section 3.4.

The participants can have four alternative navigational aids, as described below:

3.2.1 Group 1. Classical Navigation

In this group, participants are given no navigational aid. They are required to navigate to the given target location by utilizing the floor plans and other physical aids that already exist in their environments.

3.2.2 Group 2. Visual Augmented Reality Navigation

In this group, participants are given an AR headset that displays a visual path to the target location for them to follow. The participants are informed that this path leads them to the given target location. The path is displayed as a holographic visual. The participants are permitted to utilize the navigational aids in their physical environment if they so wish.

3.2.3 Group 3. Audio Augmented Reality Navigation

In this group, participants are given an AR headset that plays audio from the target location. This audio is played by calculating the reverberations of the environment, so as to simulate a realistic audio environment. The participants are informed that the audio source is located at the target location. The audio is played in a generic headset with wireless connection capabilities to not hinder the movement of the participant. The participants are permitted to utilize the navigational aids in their physical environment if they so wish.

3.2.4 Group 4. Visual-Audio Augmented Reality Navigation

This group merges group 2 and group 3's navigational aids. The participants are given an AR headset that plays audio from the target location, and displays a visual path to the target location for them to follow. The participants are informed that the audio source is located at the target location, and that the path leads them to the given target location. The audio is played in a generic headset, and the path is displayed as a holographic visual. The participants are permitted to utilize the navigational aids in their physical environment if they so wish.

3.3 Measures

While each participant was navigating their environment, key metrics were collected regarding their performance. They were timed on how fast they were completing their tasks, their positional coordinates were measured, and the head orientations of each participant was recorded.

The completion time metric is used to compare the task performances of each participant. It is also used to measure task completion accuracy of a given participant. For example, if a participant takes 30 minutes to complete their navigation, while the average of the task is 10, then their accuracy will be zero. Hence, to be successful, they are expected to complete each navigation task in a given time period. Otherwise, their accuracy is recorded as a failure.

The positional coordinates of a participant are used to calculate the total distance covered in a given task. It is also used to measure the amount of backtracking each participant has done to complete their task.

The head orientation of a participant is utilized to determine their audio-localization attempts. This aids in determining the amount of difficulty a participant may experience when navigating with audio.

3.4 Instruments

Instruments of this study include questionnaires given to the participants, as well as the software developed for this study. The details of these instruments are given below.

3.4.1 Questionnaires

As stated in section 3.2, each participant was given a set of questionnaires at the end of their experiment. These questionnaires are the NASA task load index (see Appendix A) and a demographic information questionnaire with an optional open ended user satisfaction questionnaire (see Appendix C).

NASA task load index is an established and commonly used set of questions for measuring the mental load of the relevant task [22]. The results of this question set were used in comparing each group on their mental demand.

The demographic information of each participant was taken. The participants were asked about their age, gender, area of study, previous AR or virtual reality experiences, any disabilities that they may have, and how familiar they are with the building. Additionally, they were asked how often they play computer games, as well as the genre of the game. They were also asked whether they have prior experience in any activities that require navigational abilities, such as hiking, orienteering, and scouting. Finally, the participants were given a set of open-ended user satisfaction questions for any extra notes they may have on the experiments.

3.4.2 AR Software Design

To conduct this study, a head mounted display was used in order to track the position of the participant within a This study utilizes a head mounted display, the Hololens 2, in order to track the position of the participant within a building. A separate computer communicates with the Hololens as the participant navigates around the building. This computer is also responsible for playing the ray-traced audio information for the participant to listen and respond to. To compute the reflections, the computer utilizes a digital reconstruction of the internal walls of the building it is conducted in. The system architecture is detailed below.

Software

Below is a list of libraries and toolkits used to make this study possible.

- Unity
- MRTK
- Steam Audio

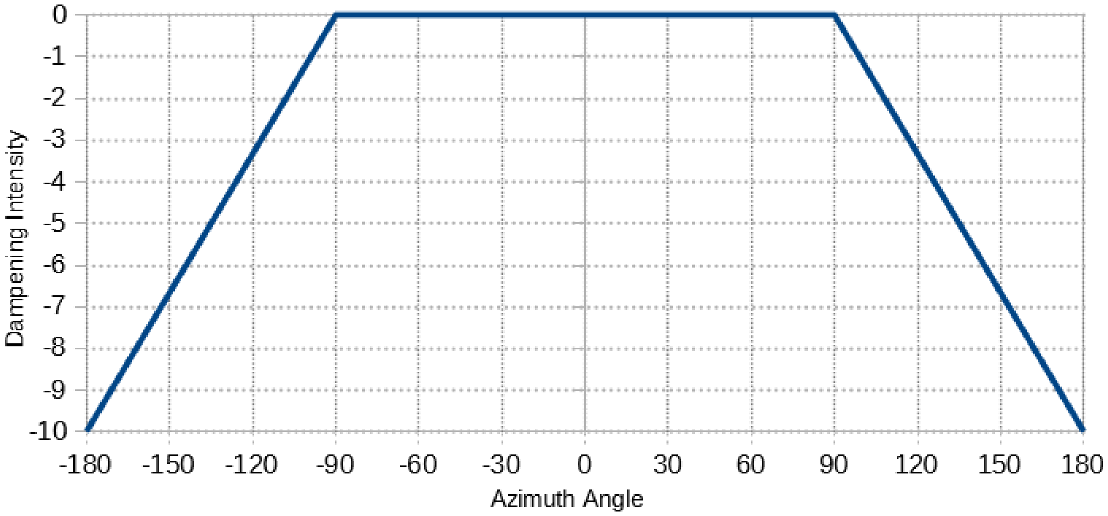
A detailed explanation of why these are used, as well as the limitations brought by these systems can be found below:

To facilitate communication with the Hololens, and interface with other external libraries, this study utilizes the Unity game engine (version 2022.3.40f). A major reason for this is due to the requirement to utilize MRTK, a toolkit which provides libraries for developing on the Hololens. MRTK is best compatible with Unity.

In addition to MRTK, this study uses the Steam audio library to generate audio that can interact and respond to the geometry of its environment. Through this library, audio can be simulated and generated in a much more realistic method than traditional methods. Due to the working nature of the Hololens, steam audio cannot be used in an embedded method native to the Hololens device. This is not a hardware limitation, but a software limitation caused by how Hololens processes audio. As a result of this limitation, it is necessary to use a separate PC to process the scene.

To minimize the amount of confusion on the front/back direction of the audio, the audio dampening technique of [24] was used. This is done by modifying the available spatial audio rendering with a low-pass filter applied to the audio a participant listens to. When the audio source is in front of the user, the intensity is left as original. However, when the audio is more than 90° away from the front of the participant, the audio source is dampened. This dampening effect is linearly interpolated between 0 decibel dampening at 90°, meaning that the audio is unchanged, to -10 decibel dampening at 180° from the front of the participant. A dampening profile showing this effect can additionally be found in Figure 3.2.

Figure 3.2 Dampening profile



This was only done in the scenario wherein the participant had a direct line of sight to the audio source. If they had a wall in between them and the source, the dampening effect would not be applied. This was done to ensure that the users would not be confused when attempting to localize the sound direction, facing away from the source, and thinking the dampening was happening due to the direction of the source. This process can be further optimized by acquiring the result of the ray-traced audio as well as the location it reflected from and dampening this result based on whether it is in front of or behind the participant. Unfortunately, due to current technical limitations, this was not possible to do.

Hardware

Below is a list of the hardware used in this study.

- Microsoft Hololens 2

- Logitech G733 Headphones
- A laptop PC

A detailed explanation of why these technologies are used can be found below.

This study utilizes the Microsoft Hololens 2 due to its mixed reality capabilities. The headset utilizes a transparent display, allowing participants to be able to see their environments clearly. If the device is turned off, the sight of the participant is not blocked. Additionally, as the visual image is not a reconstructed feed of external cameras similar, as with the Meta Quest, there is no smearing or visual error problems associated with the headset. The headset is also able to provide highly accurate 6DOF data. These advantages ensure a continuous and unmodified testing experience for the participants.

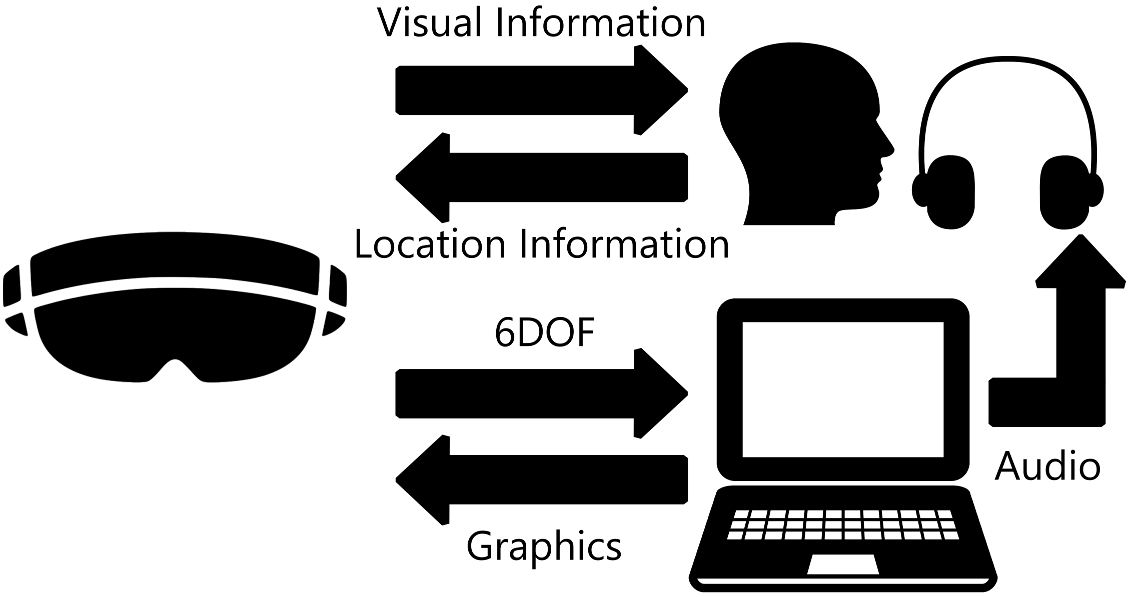
As stated in the section 3.4.2, there is a requirement to use a separate PC to process the scene the participant navigates through. Since the headset can live stream the 6DOF movement of the participant in the scene to other devices, it is possible to use a computer to acquire the 6DOF data, process what the participant should see visually and hear through audio, and then display the computed results back to the participant through the headset.

Finally, this study uses a seperate audio headset, the Logitech G733. This headset is a bluetooth wireless headset, capable of playing seamless and lossless audio to the participants. By using this system, we can allow for the participants to walk with the Hololens and hear accurate audio while not being inconvenienced by any cables, allowing for a more natural experience within the headset.

Technical Implementation

In the experimental implementation, the systems listed in the infrastructure section are connected and rely heavily on the resources provided by the PC. The information flow from each component to each other component can be seen in Figure 3.3.

Figure 3.3 Technical model



The Hololens may display to the participant a path for them to follow. When the participant moves in the scene to respond to this information, the 6DOF data is tracked by the Hololens. The data is streamed to the PC shortly after. With the location information of the participant, the PC calculates what they should be seeing at that moment, and streams the graphics information back to the Hololens. The PC also calculates what audio the participant should be hearing at their current location, and streams the audio to the headset on the participant.

3.5 Research Procedure

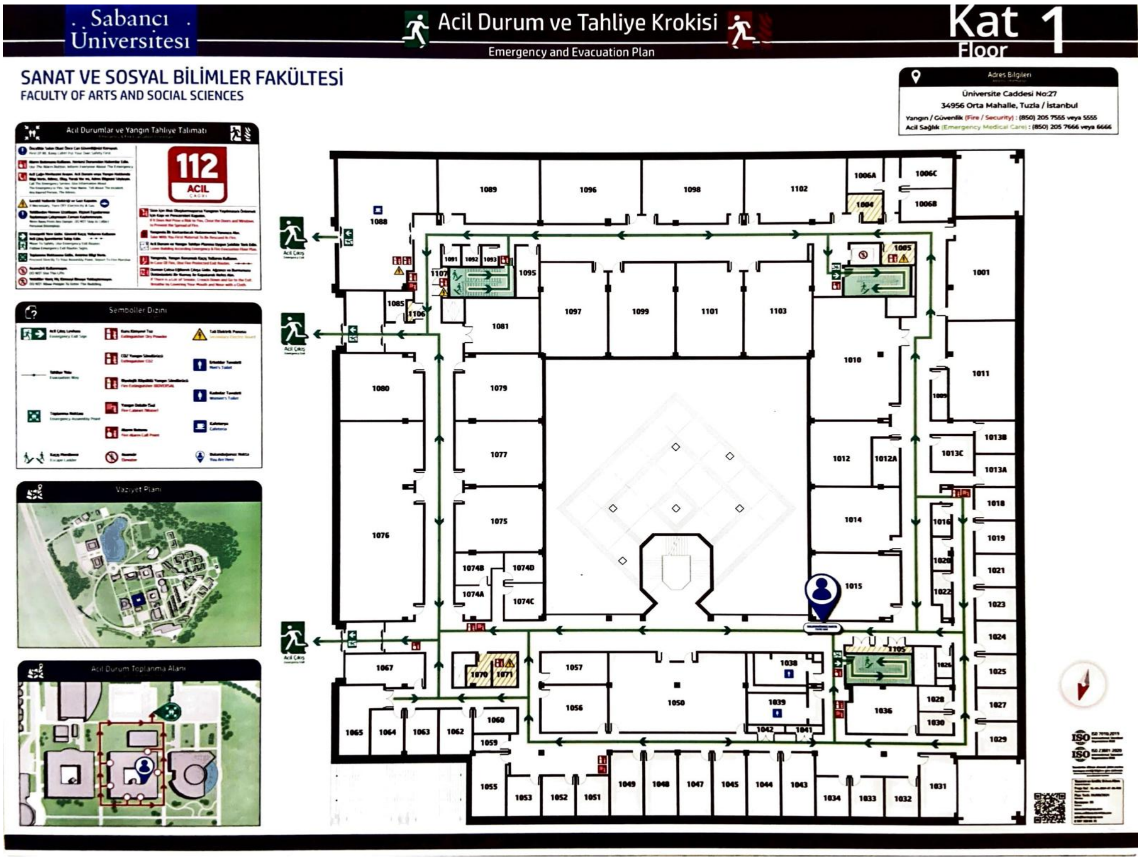
To conduct this study, an accurate geometric reconstruction of the building needs to be created. This digital twin of the building is required for the purposes of computing sound reflections off of the walls of the environment.

3.5.1 Digital Twin Creation

While the sound guidance process can be achieved without an initial model, where the internal structure of the building is generated as the participant explores the space, this results in a poorer overall experience. As the structure is constructed in partial segments, the sound simulation is likely to not be accurate. This inaccuracy can in turn lead the participant to hear sounds from the wrong directions. Similarly, without knowing the full structure of the building, creating a correct visual path indicator (Group II, Group IV) is not possible to do. Due to these reasons, coupled with the navigation task of this study, live reconstruction of the given building is not a viable solution. As a result, the digital model of the building is required before the participant starts exploring the multi-floor structure.

The digital twin creation process consisted of three steps. First, as a reference, the floor plans of the building's various floors were digitized and imported into the Unity scene. One level of these floor plans can be seen in Figure 3.4. This floor plan is the "emergency and evacuation plan," which is publicly accessible on all levels of the faculty building. As these floor plans were updated in the past three years, they are up to date with the internal structure of the faculty building. Any changes since the creation of these plans and the study were fixed within the following steps of the reconstruction process.

Figure 3.4 FASS emergency and evacuation plan



In the second step of the reconstruction process, a digital scan of the internal structure of the building was taken with the Hololens. By default, the Hololens creates a 3D mesh of its immediate surroundings and keeps them in its memory. By accessing the Hololens' online interface portal, this mesh reconstruction can be installed as a .obj file format. The .obj file format is an industry standard format for sharing 3D objects. It allows for reconstructing surfaces and meshes from a set of given points, or vertices. Unfortunately, the online portal does not make it possible to import the entire building map at once. To optimize the process due to hardware limitations, the Hololens can only access the closest 126 blocks, which approximately corresponds to a radius of 10 meters around the participant if there are no floors mapped or associated above or below the participant. If there are floors above or below the participant, and the Hololens has associated these locations with the position of the participant, it attempts to load these floors as well, reducing the accessible mapped structure mesh radius around the participant.

Due to this limitation, the digital scan was acquired in two steps. First, to load the entire internal structure into the Hololens, an initial pass of the entire building was done. While making sure that no large unmapped areas were left, a researcher slowly walked around the entire building. After the whole building interior was mapped, a second pass was done to download each patch of mapped surface. As stated previously, this only gives access to the closest 126 blocks. Due to this, more than 50 individual segments were generated and downloaded as .obj files. Each file had a portion of its segment intersecting with a previous segment. This was done to ensure that each segment could easily be tiled, allowing for an easier and more accurate reconstruction process.

In the third and final step, the the various .obj files were matched to get a rough approximation for the size of the building. From this, the floor plans were placed into the scene and matched onto the interior meshes. This gave the correct scales for the images, and allowed the meshes to be matched while taking the floor plan as the reference. Due to the interior scan containing too many vertices, and containing gaps in some walls and in the locations of windows, it was necessary to simplify the overall interior geometry to minimize the impact on performance as well as to provide definitive solid shapes with no gaps in geometry.

This process results in a relatively accurate and recent model of the interior structure of a building. By knowing the overall 3D shape and structure of a building, the acoustics of the building can then be modeled by simulating the sound wave propagation in the environment. A fairly common method of modeling the acoustic characteristics of an environment [59]. The screenshots of matched hololens .obj file outputs, as well as the resulting cleaned up version of the environment can be seen in Figure 3.5 and Figure 3.6 respectively. Additional before and after shots of the internal structure can be found in Figure 3.7.

Figure 3.5 FASS raw internal scan

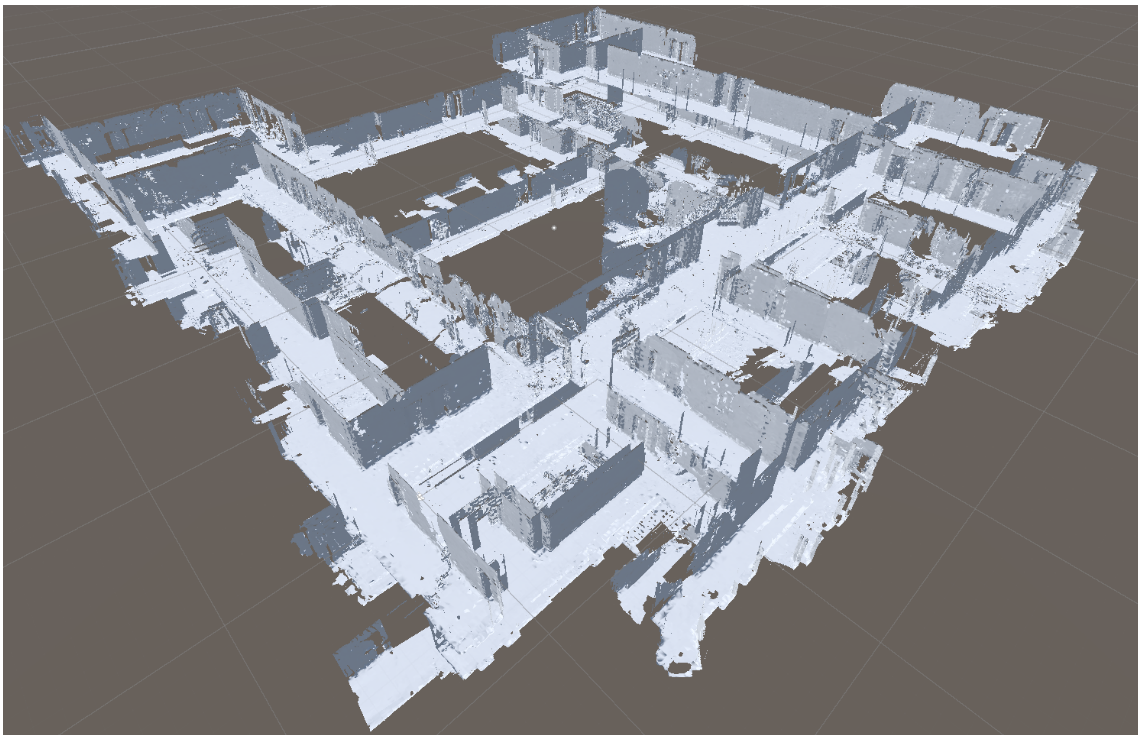


Figure 3.6 FASS processed internal scan

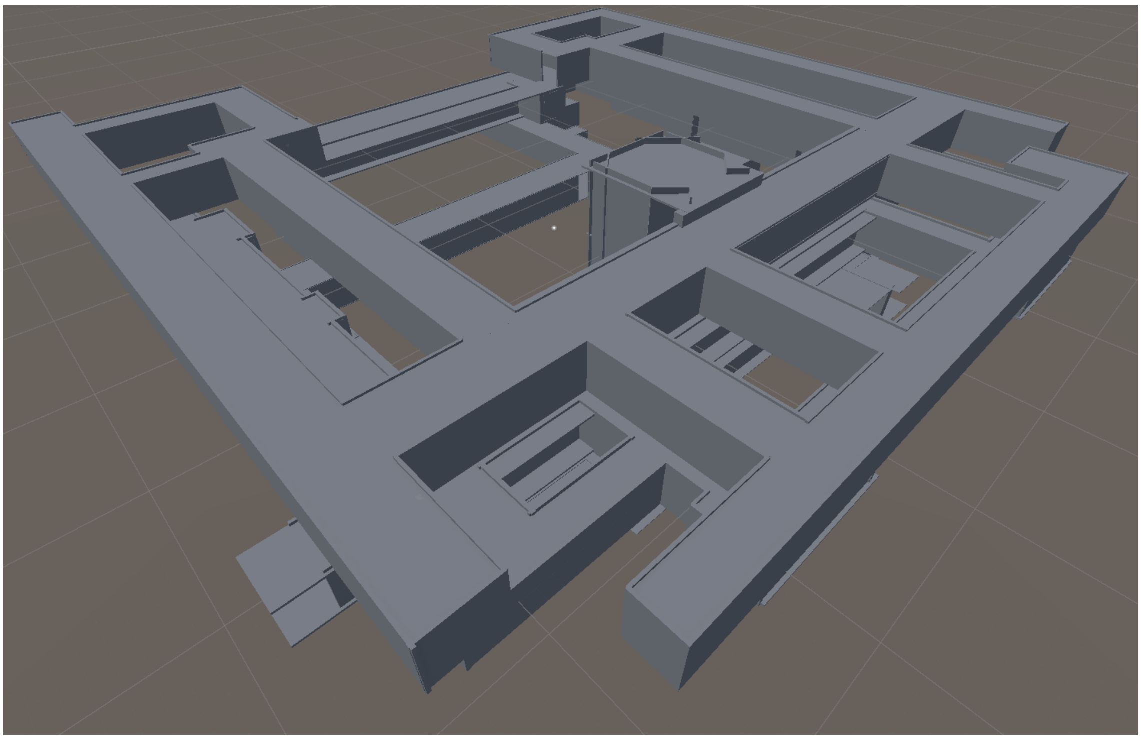
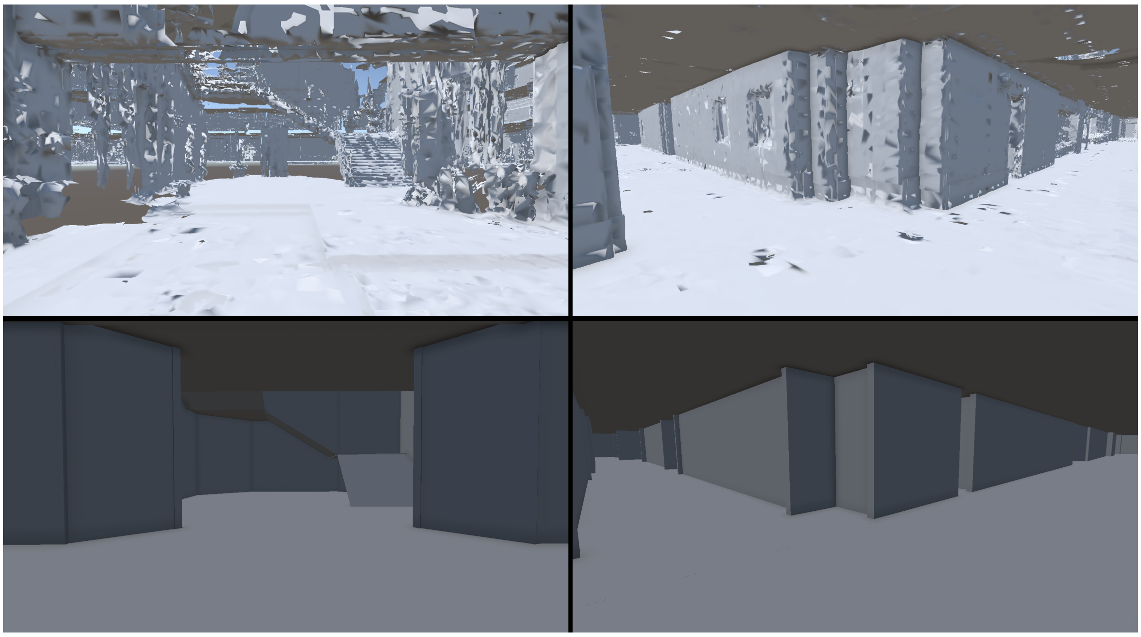


Figure 3.7 Additional FASS internal scan comparisons



3.5.2 Experiment Audio

Due to a lack of clear guidelines on the most effective audio type for a given set of tasks, it is difficult to pick an audio and know that it will work. Previous studies have shown that as long as the audio is an earcon and not a speech audio, the cognitive load of the audio tasks will be low. As shown by [54], as long as the earcon is distinct the type of earcon is not very important. The user experience improves if what the earcon represents is conceptually similar to the audio type. Because of these reasons, several alternative audio files were tested in the design of the system. These were continuous sounds, such as pink noise and sine waves, which can help with elevation perception [34], and distinct sounds, such as sonar pings, which could be conceptually associated with the search task of the study.

Distinct sounds were preferred in the design phase, as it was easier to distinguish their directions. This is due to the sonar ping being perceived as a discrete event, which can grab the attention of the user more easily compared to continuous sound events [27] and giving the user a period to adjust in between the silent phases. This type of sound can also reduce the overall cognitive load [31].

Determining what kind of distinct audio to play is a harder task. However, as there is only one audio that will play during a given task, the audio type should not affect the overall performance of the user; it will always be distinct. As a result, a free-to-use sonar sound file [26] was acquired from pixabay [10]. Pixabay is a website where users can upload royalty-free images and short sound files for other people to use.

3.5.3 Experiment Steps

The study was conducted following the steps below.

1. The participant will be asked to fill the Demographic information questionnaire (See Appendix C).
2. The participant will be assigned to one of the groups according to their entrance order to the study.
3. A preparation task will be given to the participant to help them to learn the experimental interfaces and environment. Here they will navigate to the dummy location.
4. The experiment will start.
   - 4.1. Participants will navigate to a given target location and utilize the navigational aids they are given access to.
   - 4.2. At predetermined locations, the participant will be asked whether they have seen an object in their environment (See Appendix B). This will be repeated four times.
   - 4.3. 4.1.-4.2. will be repeated four times for each target location.
5. The participant will be asked to fill NASA task load index (See Appendix A).

3.6 Data Collection and Preparation

In this study, there are three types of quantitative data to collect:

- Noticed objects: Which objects each participant noticed during their navigation tasks,
- Duration: the time it took to complete their navigation task,

- Distance: the distance they covered to complete their navigation

3.6.1 Noticed Objects

The study featured four main objects on the navigation paths of the participants, with a backup object in case they accidentally went in an unexpected direction. These four objects are a white stand, a tall coat hanger, an orange backpack, and a yellow trash can with a caricatured penguin face on its side.

When the participant passed by any of these objects, they were stopped after three seconds. If they attempted to look around, they were asked to look straight ahead. They were then asked whether or not they noticed the relevant object or not, and their answer was recorded. To control for false positives, they were also stopped and asked for objects that were not on their paths during their navigation.

An image of each object during the navigation of a participant can be found below. Figures 3.8, 3.9, 3.10, and 3.11 show the participant passing by the backpack, the coat hanger, the yellow trash can, and the white stand respectively. Furthermore, the questioning step for whether the participant has noticed the object can be seen in Figure 3.12.

Figure 3.8 Participant is passing by the orange backpack

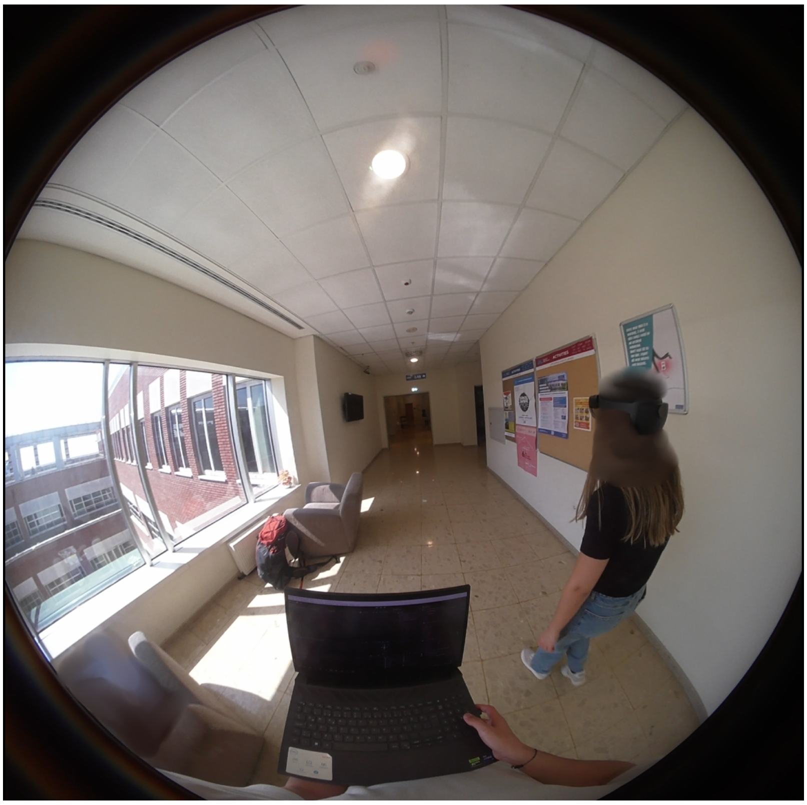


Figure 3.9 Participant is passing by the coat hanger

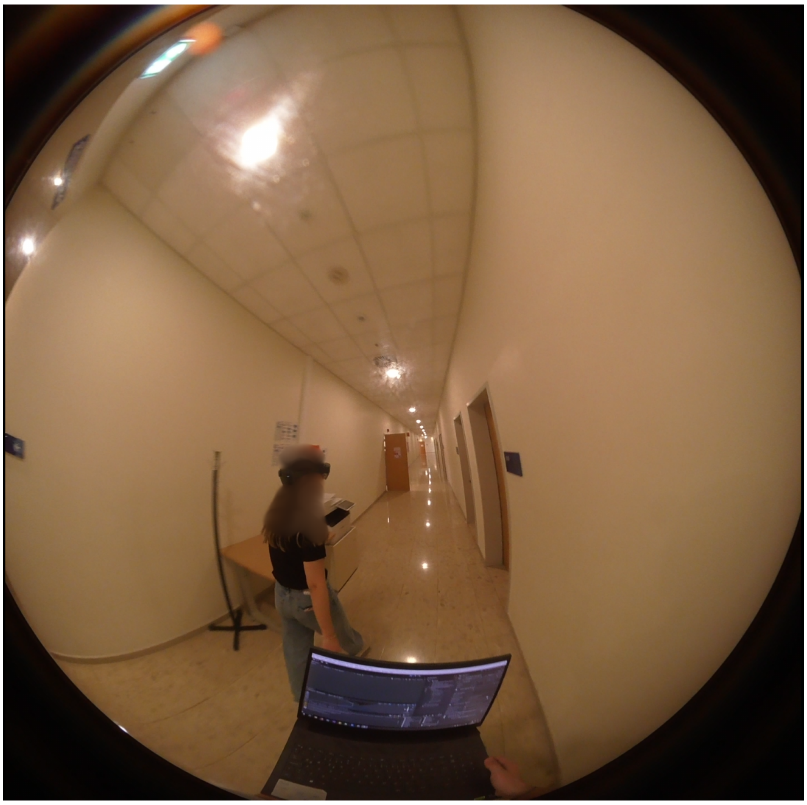


Figure 3.10 Participant is passing by the yellow trash can

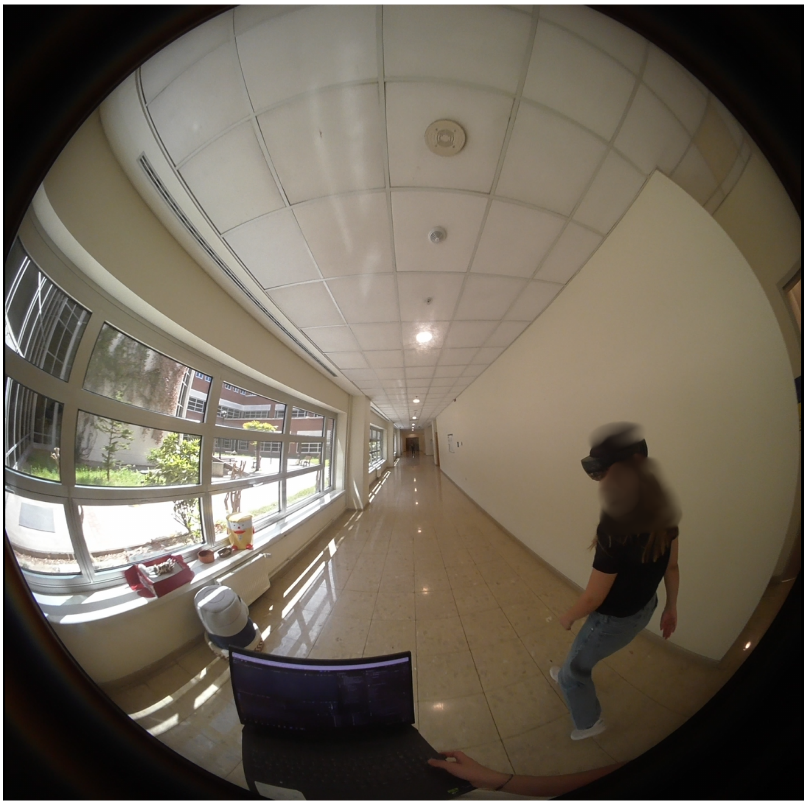


Figure 3.11 Participant is passing by the white stand

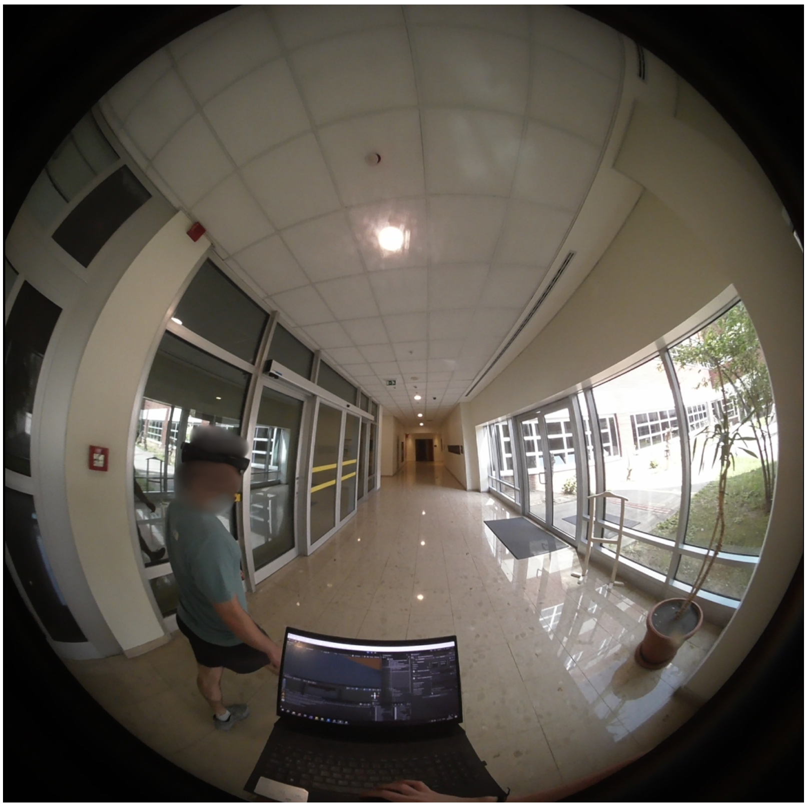
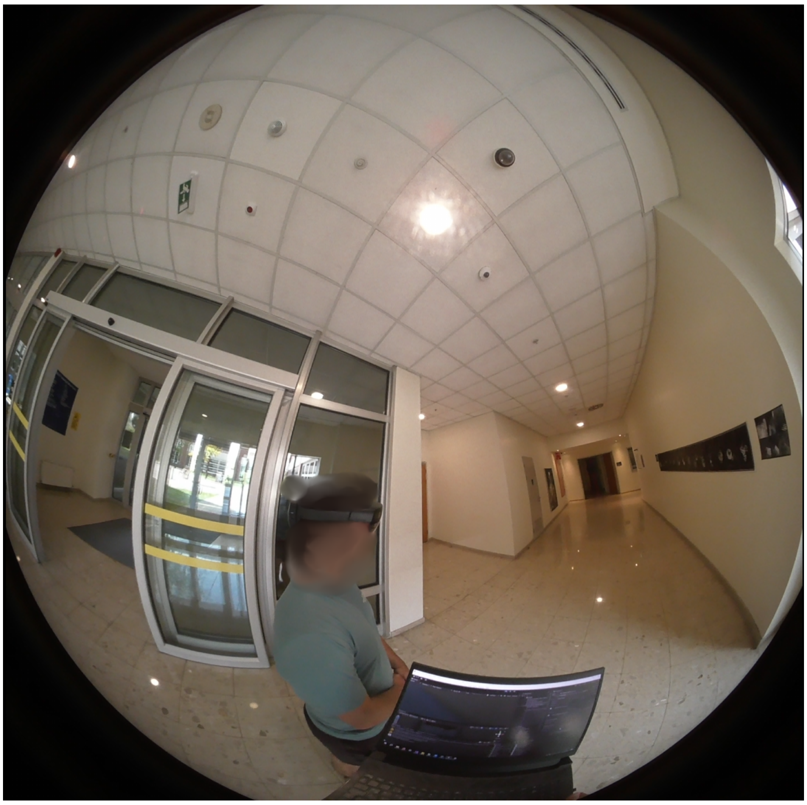


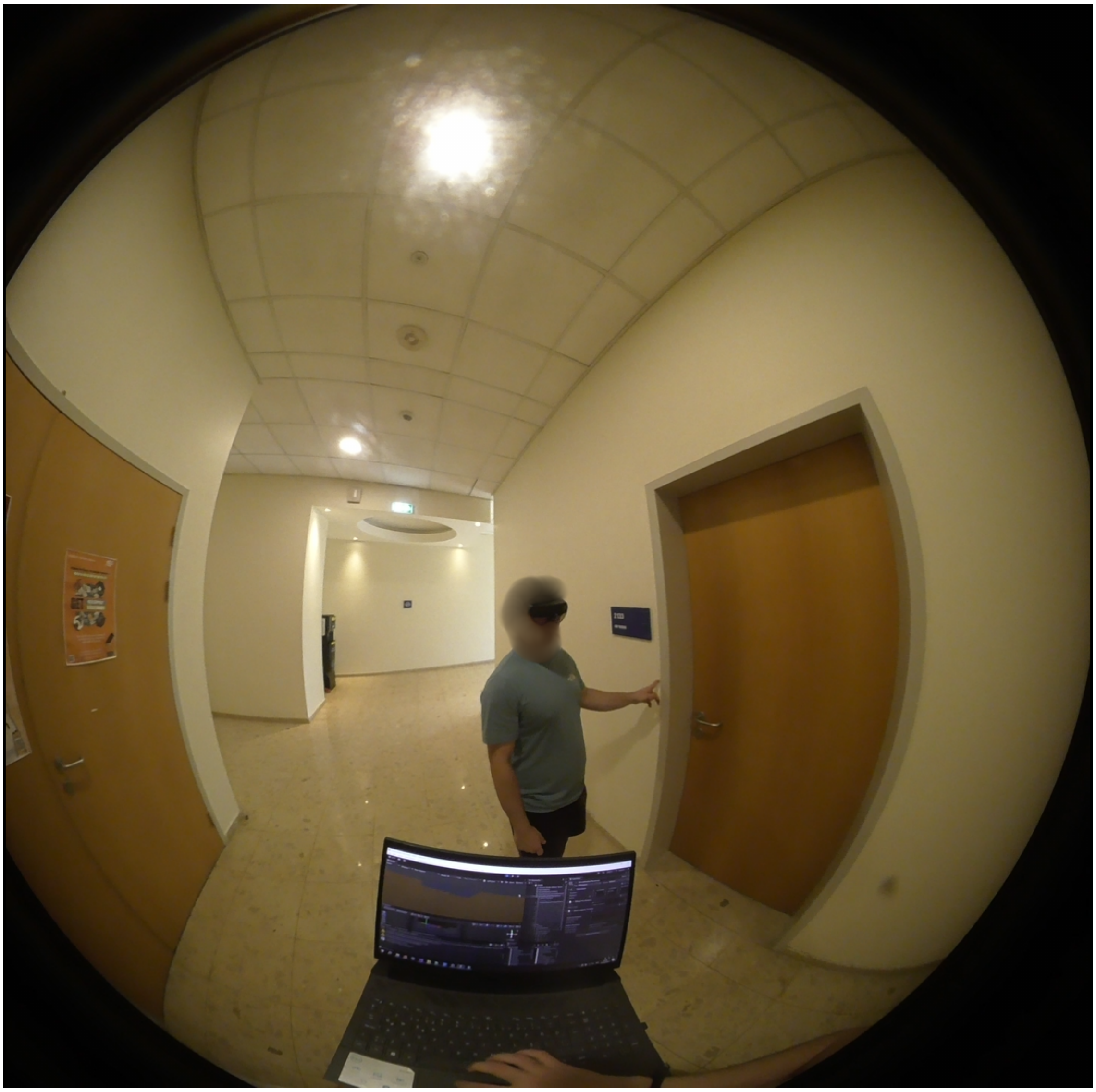
Figure 3.12 Participant is stopped to ask whether they saw the white stand



3.6.2 6DOF Data Collection - Time-Distance Calculations

The tracked data was recorded with respect to the participant's head, so for brevity the 6DOF data of a participant's head will be referred to as the participant's own data. To find the time and distance results, the position of each participant was tracked during their navigation until they reached their target destination. An image of a participant reaching their target can be seen in Figure 3.13. Every time step, which roughly corresponds to a frequency of 0.01 seconds, the 6DOF data of each participant was recorded. Due to dynamic load on the system, this reporting time can vary by around 0.005 seconds.

Figure 3.13 Participant reaching the target location



To ensure minimal errors, two methods of time keeping were used. The first method was the amount of seconds that has passed since the start of the task for this time step. This data was taken directly from the Unity system, with the use of the "Time.time" method. The second method was the UNIX time of the same time step with the use of the "DateTimeOffset.Now.ToUnixTimeMilliseconds()" function. Both of these time steps were recorded with millisecond precision to minimize the chance of data loss due to imprecision. An example showcasing how these datapoints were recorded can be found in Table 3.2.

As seen in these examples, the reported time metrics of a task is 0 at the start of a task. They are incremented as often as possible, which then corresponds to a frequency of roughly 0.01 seconds. The milliseconds precision UNIX time of the

Table 3.2 Example time data

App time	UNIX time	other 6DOF data
0	1742826044037	...
0.0199999995529652	1742826044933	...
0.0811130031943321	1742826044974	...

same time metric is also reported alongside it.

UNIX time, also known as epoch time, is a standard method of time keeping in all modern computing systems, and is one of the most reliable methods of getting the time of a system. It is defined as "a value that approximates the number of seconds that have elapsed since the Epoch" [51]. The epoch is defined as 00:00:00 UTC of 01/01/1970.

The 6DOF data consists of three types of data: the position of the participant in the digital space, the Euler rotation of the participant, and the quaternion rotation of the participant. Both the Euler rotation and the quaternion rotation describe the head rotation of the participant. This metric is reported in two different formats to ensure minimized data errors in analysis, similar to the time metric.

The position data is reported in three parts: the x position of the participant, the y position of the participant, and that z position of the participant. The Euler rotation is reported with the rotation of the participant with respect to the the x, y, and z axis. Finally, this same rotation data is reported in the quaternion format; with respect to the the x, y, z, and w axis. An example of these data points can be found in Table 3.3.

Table 3.3 Example 6DOF data

Time records	position x	position y	position z
...	0	0	0
...	-0.00628691259771585	0.0665859878063202	0.114001110196114
...	-0.00619979901239276	0.066442146897316	0.114384643733501
rotation (Euler) x	rotation (Euler) y	rotation (Euler) z	
0	0	0	
1.46667635440826	359.484832763672	358.865539550781	
1.5352908372879	359.451721191406	358.865631103516	
rotation (Quaternion) x	rotation (Quaternion) y	rotation (Quaternion) z	rotation (Quaternion) w
0	0	0	1
0.0128425629809499	-0.00436850357800722	-0.00984141416847706	0.999859571456909
0.0134440846741199	-0.00465120794251561	-0.00983385089784861	0.999850511550903

Due to the number of metrics reported, the example is broken into three rows. Each row in the table corresponds to a subsequent time point. As seen in these examples, the world position, the Euler rotation, and the quaternion rotation of the participant at each time step is reported by the system.

The time metric of each reported step was used to compute the duration of each task. This was done by finding the time difference between two adjacent valid steps, and taking the sum of all of these differences. How a data point was determined to be valid is explained later in section 3.6.3.

Similarly, the distance covered by a participant in a task was computed by calculating the Euclidean distance between two adjacent valid steps. This can be found by taking the difference of each position metric, squaring each result, finding the total sum of these squared results, and taking the square root of the final summation.

Table 3.4 Example data for time/distance calculation

Column	Time	position x	position y	position z	Valid
1	0	0	0	0	0
2	0.01999	0.00021	-0.00027	-0.00021	1
3	0.08011	0.00035	-0.00064	-0.00005	1
4	0.11598	0.00045	-0.00037	-0.00096	1
5	0.13599	0.00037	-0.00051	-0.00110	0
6	0.15168	0.00042	-0.00039	-0.00136	1
7	0.16791	0.00065	-0.00052	-0.00132	1
8	0.18928	0.00039	-0.00055	-0.00137	0

For example, for a sample data as in Table 3.4, the time and distance metrics are calculated as follows. The complete code implementation of the following steps can also be found in Appendix F.

1. All adjacent rows with "Valid" value of 1 are grouped together.
   - In this example, the rows 2, 3, 4 would be one group, and 6, 7 would be another group.
2. Each adjacent row in a given group is checked from earliest to latest. The elapsed time of a group and the total distance are calculated.
   - 2.1. Total time is initially 0. With every adjacent valid row, the time difference between the rows is calculated with $\Delta T = Time[row] - Time[row-1]$.
   - 2.2. Total time is updated by adding ΔT to total time: $TotalTime = TotalTime + \Delta T$
   - 2.3. Total distance is initially 0. With every adjacent valid row, the distance between the rows is calculated with Euclidian distance: $\Delta D = \sqrt{(x_{row} - x_{row-1})^2 + (y_{row} - y_{row-1})^2 + (z_{row} - z_{row-1})^2}$

2.4. Total distance is updated by adding ΔD to total distance: $TotalDistance = TotalDistance + \Delta D$

Following these steps, the data from Table 3.4 would be processed in the following way. First, the group of 2, 3, 4 and 6, 7 would be created. Next, the group of 2, 3, 4 would be processed for time and distance; the time difference of 2-3 would be calculated as 0.06012 and added to total time. The time difference of 3-4 would be calculated as 0.03587 and added to total time.

Similarly, the Euclidean distance between these row pairs would be calculated as 0.000426732 and 0.000954463 respectively, and added to the total distance. Following this, the grouping of 6, 7 would be processed. The time and distance would be calculated as 0.01623 and 0.000267208 respectively. At the end, with no other groups left to process, the total time and total distance would be reported as 0.11222 and 0.001648403 respectively.

3.6.3 Data Cleaning Process

Before computing the final time and distance results of a task, the data was cleaned by reviewing the movements of the participant. Due to the infrared tracking technology of the Hololens, and the connection of the Hololens to a PC over WiFi, there are certain situations where tracking can be lost or lose tracking quality. In such situations, the position and rotation coordinates received from the Hololens can be wrong by orders of magnitude. In addition, the time period in which the participant is answering questions should also not be processed, as this does not relate to their performance of the task itself.

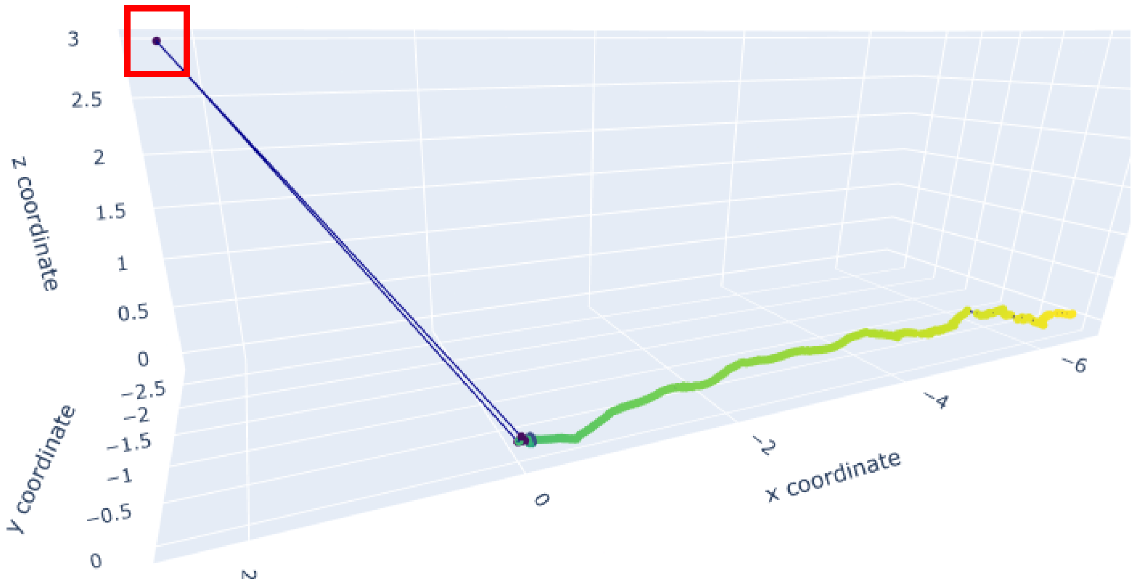
To ensure the capture of parts of the data that should not be processed, the researchers marked on the data whether the participant was stopped or not, dubbed as "valid". An example of this can be seen in Table 3.4. Whenever a participant was stopped, either for questioning or for recalibration purposes, this validity check was updated accordingly. This allowed the researchers to be aware of whether or not a participant stopped due to external factors or due to them looking around at information sources such as floor plans.

Although the sections where the participant is stopped by researchers are marked, the sections which are wrong due to brief tracking problems are not marked immediately. So, to ensure an accurate reflection of the valid and non-valid markings, the movement results of the participants were reviewed by the researchers. The data sections with errors were marked as non-valid as conservatively as possible to ensure that the time and distance results were reflective of the real performance of the participant.

For this purpose, a python script was written in Blender. Blender is a modeling software which supports easy scripting and generation of connected vertices. It has a 3D scene editor, which makes it easy to view the imported path of a user. The code can be found in Appendix E. This code extracts the position coordinates of the given task file, creates a mesh from each of these vertices with edges connecting adjacent coordinates, and renders the index of each vertex on top of them.

Due to tracking errors, the tracking data can contain dirty data. An example of one such data can be seen in Figure 3.14. The section that is boxed in with a red square is the section which is considered to be dirty data. These sections are marked as non-valid. In this task, the participant is going towards the negative x direction, where the gradient represents the movement order. The participants starts the task at origin, shown as dark markers, and the markers get lighter as they move forward.

Figure 3.14 Example error data, highlighted by the red box. Gradient of points show travel direction, going from darker to lighter



3.6.4 Data Integration Process

Following the cleaning and final performance calculation, the results of participants for each of their tasks were saved into a spreadsheet. Along with analysis results for time and distance, the spreadsheet also contains the group the individual belongs to, the objects they noticed during their tasks, and their questionnaire answers. Along with the questionnaire answers, the time when the participants submitted their answers is also recorded. Each row in this spreadsheet represents a unique participant's data. The final data headers in this spreadsheet can be found in Table 3.5.

Table 3.5 Integrated data headers

Header	Explanation
ID	participant identifier
time stamp	when the questionnaire was answered
Group	experiment group of the participant
Object 1	if they noticed the backpack
Object 2	if they noticed the white stand
Object 3	if they noticed the trash can
Object 4	if they noticed the coat hanger
Task 1 time	measured in seconds
Task 2 time	measured in seconds
Task 3 time	measured in seconds
Task 4 time	measured in seconds
Task 1 distance	measured in meters
Task 2 distance	measured in meters
Task 3 distance	measured in meters
Task 4 distance	measured in meters
Age	see Appendix C Q2
Gender	see Appendix C Q3
Faculty	see Appendix C Q4
XR Exprience	see Appendix C Q5
Env. Familiarity	see Appendix C Q6
Nav. Skill	see Appendix C Q7
Disability	see Appendix C Q8
Gaming Frequency	see Appendix C Q9
Games Genre	see Appendix C Q10
Nav. Activities	see Appendix C Q11
Experience Notes	see Appendix C Q12
Extra Notes?	see Appendix C Q13
Mental Demand	see Appendix A Q1
Physical Demand	see Appendix A Q2
Temporal Demand	see Appendix A Q3
Performance	see Appendix A Q4
Effort	see Appendix A Q5
Frustration	see Appendix A Q6

The task time data columns are measured in seconds, while the task distance data columns are measured in meters. Furthermore, the columns labeled with "Object 1," "Task 1 time," and "Task 1 distance" all belong to the same task, task 1, meaning they are the results of the participants performance on the same path. Similarly, the other numbered headers all belong to the other three tasks.

To analyze the overall results of each group, the participants in this spreadsheet were sorted and separated into their own groups. As a result, four sections emerged, each belonging to one experiment group. To evaluate the overall time and distance performance of each group, the calculated results of every individual were averaged for each task. So for each experiment group, a total of eight average results were generated, two from each task. Additionally, the total number of objects noticed for each group was calculated and normalized against the number of participants in that group. This generated one more result for each group.

3.7 Data Analysis

As four groups of people were compared with each other in this study, an ANOVA analysis was necessary. The average performances of each group was computed, and the groups were compared on this average. The groups were compared on their task completion performance, comparing how fast the groups completed their respective tasks. Additionally, the groups were compared based on their environmental awareness by comparing how many objects they were able to identify successfully vs how many false positives they reported. The groups were also compared in the amount of mental load they have experienced during their tasks.

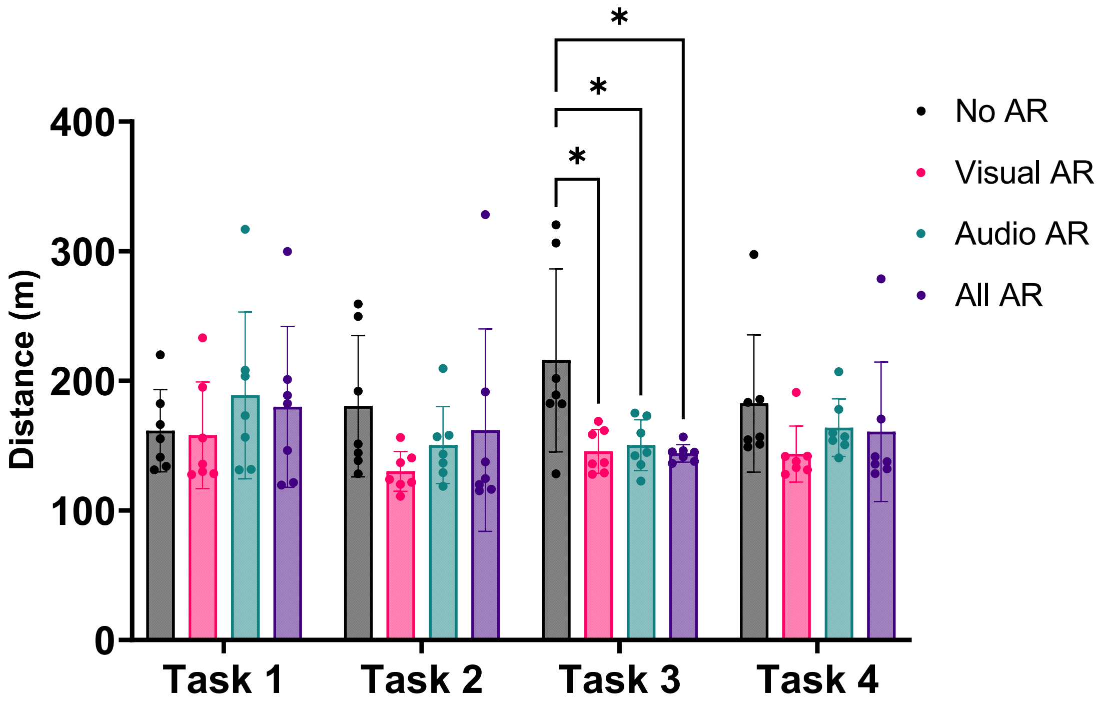
4. RESULTS

In this study, to answer each research question, both qualitative and quantitative data was collected. Quantitative data was analyzed by the use of statistical methods, while the qualitative data was analyzed to find supporting evidence for the statistical results. The statistical analyses presented in the following sections were computed by using GraphPad Prism [20], a statistical data analysis application made for scientific research. The results of these analyses are given below, such that each section answers one research question.

4.1 Navigation Performance Analysis

The navigational performance of each participant were evaluated based on their task completion times, as well as the distance they have covered during their navigation. The lower these metrics, the better they were at navigating their environment. These metrics were compiled and grouped for each testing group. The distance results were measured in meters, and the time results were measured in seconds. The analysis results for distance performance can be found in Figure 4.1, while the results for time performance can be found in Figure 4.2. Both the distance and time performances of the participants were analyzed and compared with 2-way ANOVA. Any statistically significant differences between each group in these figures are marked with a * icon and connected to each other. Additionally, the group referred to as "all AR" is the group who utilized audio and visual AR.

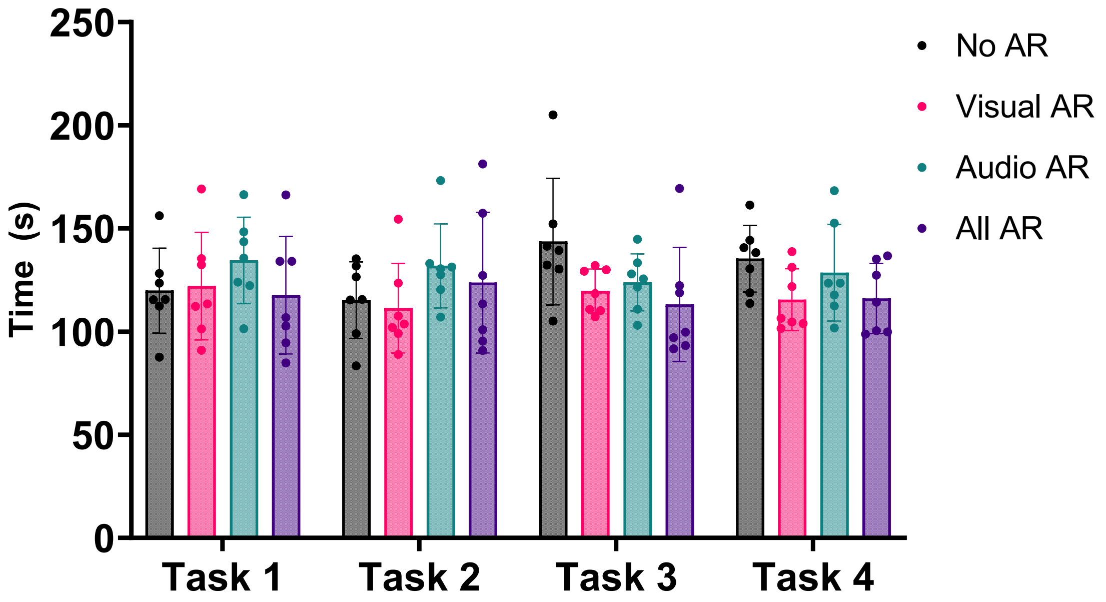
Figure 4.1 Distance Analysis Results



Here, it can be seen that the covered distance results show a statistical significance specifically in task 3. This indicates that in task 3, the participants who utilized AR technologies were significantly better at completing their tasks in the least amount of distance covered. Although there is no statistical significance observed in tasks 2 and 4, the participants' distance performance when utilizing AR technologies showed a better result on average in comparison to normal users.

In contrast to the distance results of tasks 2, 3, and 4, task 1 did not show a better distance performance for AR users. In task 1, it is observed that no AR users had an average performance similar to that of visual AR users, while the performances of audio AR and all AR users were worse on average.

Figure 4.2 Time Analysis Results



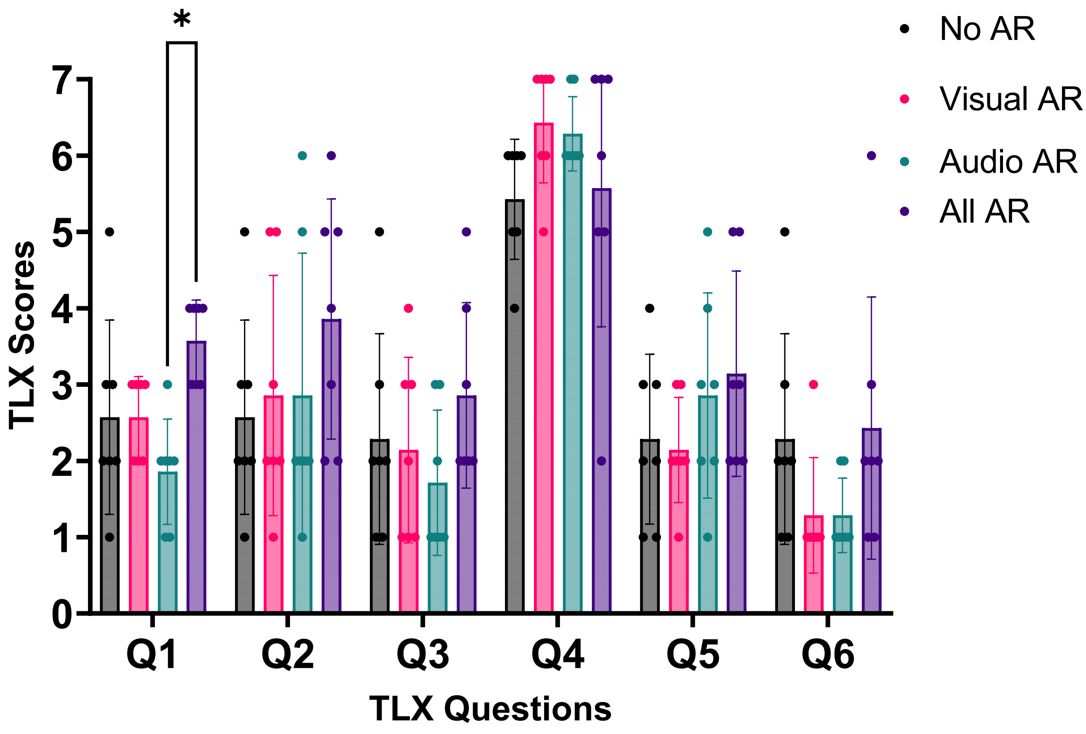
Analysis of the time performances of the participants did not show any statistically significant difference among each group. Although there are no significant differences, it can be observed that, in tasks 3 and 4, the groups who utilized AR technologies had a better average time performance in comparison to those who did not utilize AR. However, this same difference pattern is not observed with tasks 1 and 2. In these tasks, the worst average performing group was those who used audio AR.

These results reflect the qualitative feedback of the participants as well. Here, participants will be referred to as P, followed by their identification number. So if we are referring to participant with an identification number of 5, they would be referred to as P5. During their test, P31 utilized visual AR assistance. At the end of the test, they stated that without any visual directives, they would have been very lost. A similar sentiment was also reflected by P20, who was given the audio and visual AR hybrid system. They stated that when technical issues happened with the visual AR system, such as positional drift, the audio system was a good fallback to rely on. Finally, P18, who utilized audio AR, stated that audio was very useful in understanding whether or not they were getting closer or further from their target location. Additionally, they stated that it was useful in affirming them in whether or not they were going in the correct direction.

4.2 NASA TLX Results

Each participant was asked to give their answers to the NASA task load index. The analysis of these answers were done by mapping the answers of each group with respect to each question. The differences between each groups answers were compared using 2-way ANOVA. The analysis results for the task load index can be found in Figure 4.3.

Figure 4.3 Answers to NASA TLX questions



The TLX questions specified in the figure can be found in Appendix A, where Q1 refers to the first question in this appendix, Q2 refers to the second, and so forth. Any statistically significant differences between each group in this figure are marked with a * icon and connected to each other. Additionally, the group referred to as "all AR" is the group who utilized audio and visual AR.

In the results for Q1 (How mentally demanding was the task?), there is a statistically significant difference in the perceived mental load experienced by users who utilize audio AR versus users who utilize all AR. Furthermore, audio AR users reported that they experienced the lowest perceived mental load on average of all groups. The average mental load experienced by those who used no AR and those who used visual AR was similar. In Q2 (How physically demanding was the task?), the answers of the participants show that those who used all AR perceived the highest amount of physical demand on average, while the other groups reported a lower perceived demand on average. The perceived physical load of audio AR users and visual AR users were similar on average, while slightly higher than no AR users.

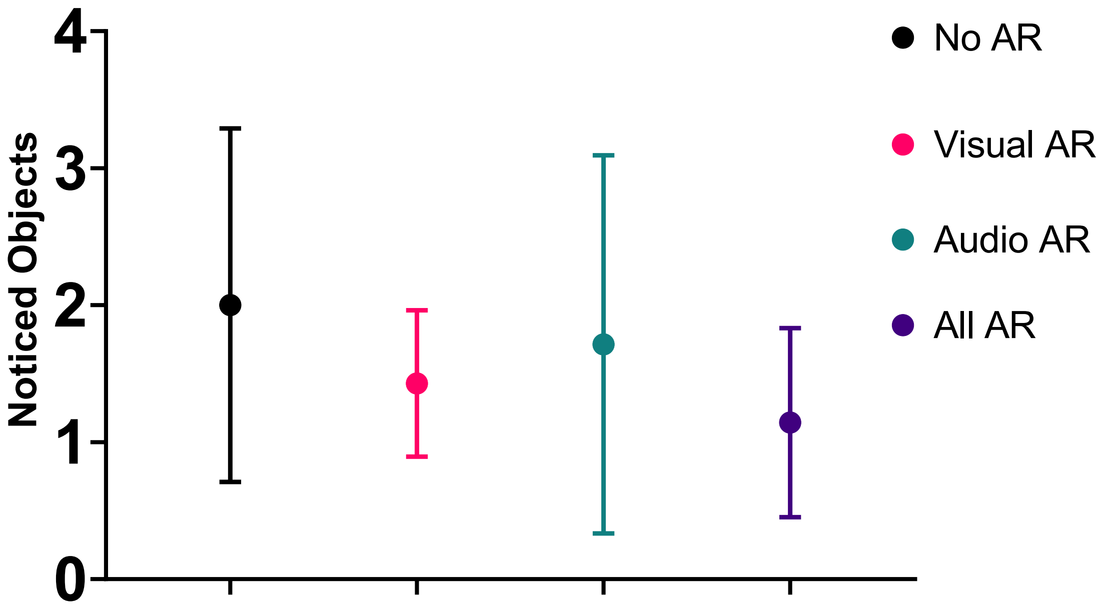
In Q3 (How hurried or rushed was the pace of the task?), the results show that the participants who used audio AR felt the least rushed on average, and the participants who used all AR felt the most rushed on average. This result is in line with the results of Q1, as participants who felt more in-control and going at the test at their own pace would also be expected to have a lower cognitive load. In Q4 (How successful were you in accomplishing what you were asked to do?), all groups reported similar and high rates of successes in the completion of their given tasks. The groups who used visual AR and audio AR reported the highest average perceived successful completion. However, the participants who used all AR did not have similar answers, leading to a large confidence window. The participants who used no AR assistance on average reported a lower perceived successful completion.

In Q5 (How hard did you have to work to accomplish your level of performance?), the participants who used no AR assistance, and those who used visual AR on average reported the lowest amount of perceived effort required in the completion of their tasks. In contrast, those who used audio AR and all AR reported a higher average effort required in the completion of their tasks. The results of the audio AR group here contrast the results found in Q1 and Q3, since a lower cognitive load and a lower temporal demand would indicate that the participant would have to work less overall. Finally, in Q6 (How insecure, discouraged, irritated, stressed, and annoyed were you?), participants who used visual AR and audio AR had a similar and the least average amount of perceived frustration with their interfaces. On the other hand, those who used no AR and all AR reported a higher and similar perceived frustration.

4.3 Environmental Awareness Analysis

To analyze the environmental awareness of each participant, the number of items they noticed during their tasks were summed up. So if a participant reported on noticing the orange backpack and the white stand, but not the other two items, they would be marked as noticing two items. These total noticed items of each participant was grouped in accordance with the technology they utilized. The final result of this process can be found in Figure 4.4.

Figure 4.4 Environmental Awareness Results



Any statistically significant differences between each group in this figure are marked with a * icon and connected to each other. Additionally, the group referred to as "all AR" is the group who utilized audio and visual AR. The number of noticed objects by the participants do not show any statistically significant difference.

While no statistically significant results can be observed with the current experimental results, some observations can still be made regarding the overall performance characteristics of each group. Here, it can be seen that the average environmental awareness of the participants who used visual AR and those who used all AR were the lowest, while the average awareness of those who used audio AR and no AR were the highest. Furthermore, the results of the audio AR group, with their confidence interval and average awareness, show a similar profile to the no AR group.

4.4 Additional Qualitative Feedback

Other than the supportive feedback given in the previous sections, some participants made additional valuable comments regarding their AR interfaces during their tasks and in the post-test questionnaire. The comments by each group, as well as which participant stated this comment, are listed in this section.

4.4.1 Visual AR Feedback

Visual AR guidance is one of the most common and well researched navigation techniques in AR settings. As such, the comments made in this section reflect prior research results regarding its positives and negatives. In fact, a comment regarding both of these aspects were made by two participants using visual AR navigation. P15, who rated themselves as fairly familiar with AR/VR technologies (with a self-score of 4/5), stated at the end of their test that the "visual path was distracting" and that they "felt like [they] had to focus on it constantly even though [they] knew not to." This fixation on visual AR elements can happen with novice users of a system, while experienced users learn to tune out this information into more background information. However, it still remains as a visual obstruction.

On the other hand, P27 and P31 stated that the visual AR information was conveyed very well, and that they really liked how clearly it showed the path to the destination. They also stated that they were not familiar with the building and that they would have likely been lost easily without these directives. Both of these participants reported that they were not familiar with AR/VR technologies (with a self-score of 1/5). These comments show that for novice users, the visual AR interface was sufficiently well designed and was useful, complementing the quantitative results.

4.4.2 Audio AR Feedback

In comparison to visual AR users, who utilized the system in similar methods, audio AR users used their navigational aid in two distinct ways. Some participants, such as P11 and P22, stated that they used the audio as only a secondary assistance, and that they relied more on physical navigation aids like the floor plans. Other participants, like P18 and P26, tried to rely more on audio while keeping their reliance on physical aids to a minimum.

A common feedback between each of these participants was that sound was very useful in reinforcing whether or not they were going in the correct direction, but that it was not a strong aid in understanding where the correct direction to navigate to was in junction points. An important observation of note made by the researchers in this aspect is that audio AR users tended to instinctively pick the correct direction to go towards even though they consciously stated that audio did not help in this aspect.

Participants with no AR assistance tended to wander around after finding the correct floor until they found a floor plan. In comparison to this, audio AR users either listened to their interface for a few seconds before deciding the direction they needed to head towards, or picked a direction to go towards immediately after exiting the stairs. In all these cases, except for one exception with P22, participants were able to accurately pick the correct direction to head towards. P11 and P22, who used audio as a secondary assistance, tended to stop to look at a floor plan on their path, while participants like P18 and P26 tended to explore while following their audio guidance.

As stated previously, P22 used the audio aid as a fallback, or a secondary assistance. Due to this, when navigating the path for task 4, P22 first identified a floor plan when going in the correct direction and utilized it. When task 3 was assigned to them at a later point, they first navigated to this floor plan, found where to go, and then returned back, where they continued navigating the expected shortest path.

4.4.3 Audio and Visual AR Feedback

Unlike only audio or only visual AR users, audio and visual AR users, referred to here as "all AR," tended to use their system in their own unique ways, making broad categorizations difficult. Some participants made use of only the audio or only the visual AR aspects, ignoring the other modality entirely. Other participants tried to use both equally, attempting to cover the shortcomings of one system with the advantages of the other when necessary. The remaining groups relied mostly on one aspect of the system, either audio or visual, using the other AR aspect as a complementary tool that may be useful but not worth paying much attention to.

As an example of the users who focused only on one aspect, there is P12. This person stated that they relied only on the audio interface during their navigation, and did not pay much attention to the visuals. In contrast, P28 did not pay any attention to the audio around them. This is also likely the reason why they commented that they "did not understand the sound even came from the target destination," even though they informed and shown that this was the case at the start of their test. Similarly to P28, P32 also stated that they did not make use of their audio system. They stated that the audio was just annoying and too loud when close to the target.

Other participants, P16 and P24, made feedbacks showing that they were attempting to utilize the audio modality along with the visual modality. P16 stated that the visual aids were very helpful, and that it was what helped them the most while navigating. They further stated that audio was helpful when moving vertically between floors, but was not helpful in navigating the floor itself when far away from the target. They speculated that this might be caused by the environment being very large.

P24 made similar comments to P16, stating that the audio was useful in understanding the vertical direction when going up and down the stairs. They further stated that the audio reverberations, caused by sound simulation of the environment, made it hard to understand which direction to go when they were close to the target location. They stated that when they got too close, the audio was too overwhelming due to being too loud. As a solution, they suggested the system to be dynamic and/or tunable by the user.

Finally, P24 made some important comments regarding the ability of the audio system to convey information during their navigation task. At the start of their test, they informed the researchers that they would challenge themselves in understanding where they were supposed to go. This was their own initiative and they were not prompted by the researchers to do so. As a result, before they started their vertical navigation section of their tasks, they attempted to localize which direction they would go on the targeted floor. In all four navigation tasks, they accurately identified the direction in which they would be required to go in. That is, in task 1, they identified that the audio came from "above and to the left of the stairs." Similarly, in task 4, they identified that they would need to navigate "down and to the right of the stairs" based on the audio.

Unlike the above participants, P20 opted to focus mostly on the visual aspect of their AR interface, using the audio as a secondary assistance. They stated that they "tended to focus mostly on the visual path indicator" ahead of them, which "made it hard to pay attention to [their] surroundings." This also complements a comment they made in their post-test questionnaire, where they stated that they slightly felt dizzy when using the stairs due to the visual indicator and the stairs overlapping. This visual overlap, while covering a very small portion of the stairs, was enough to make them feel as though they were losing their balance. Finally, they stated that when they lost sight of the visual indicator, either due to instrument error causing drift or them losing track of it, the audio system was useful in tracking to the target destination or until they found the visual indicator again.

The comments made by participants in the all AR group complement the results of the participants in the visual AR group, while contrasting with some of the comments of the audio AR group. There are certain common critiques by participants in the all AR group that, if caused by the audio system, would have been commented on by the audio AR group as well.

5. DISCUSSION

The aim of this study was to understand the effectiveness of audio guided AR in complex, 3D indoor environments. To evaluate this, we proposed three questions. RQ1 was established to determine how different aspects of AR affect the time and distance performances of navigation tasks. The results indicate that in terms of distance performance, people tend to outperform non-AR users. Similarly, the time performances of AR users also tend to outperform non-AR users. In both of these metrics, visual AR users have the highest performance on average. It is shown that in their distance performances, people can significantly outperform non-AR users. While this study has shown this to be true for one test case, test 3, it is possible that the results for tasks 2 and 4 may also become statistically significant with a larger participant count. This speculation is justified by the profiles of the results in Figure 4.1, where it can be observed that the group which utilized no AR assistance consistently had the lowest average distance performance, as they traveled the longest distance. With more participants, the uncertainty around the performances of each group may reduce, leading to statistical significance in tasks 2 and 4. On the other hand, the distance result of task 1 does not show a profile similar to the other tasks.

It is possible that task 1 was unusually easy in comparison to the other tasks. If a task is too easy, we would expect the performances of all participants to be close to each other, making it harder to observe meaningful results. To determine if this is indeed the reason, the paths need to be analyzed and their difficulty ratings need to be numerically expressed. This is a difficult analysis to complete, as to accomplish this, the visual environment, the number of decision points, the complexity of the decision points and many more factors need to be considered. Literature in this area has tended to focus on complexity in maps, such as digital maps [62; 32] and physical maps [63]. In literature, there are known parameters that affect the complexity of a certain path a person can take. For instance, the path length, which is the total distance to a given destination, can be a factor. Additionally, obstacles and decision points can also affect the complexity of a path. Decision points can be any location where a person might have to decide on whether to continue forwards, turn to the side, or turn and go down a different corridor at a junction point. These are known factors, and so have been used in digital maps previously [62; 32].

While there are certain known complexity parameters which contribute to the complexity of a path [38; 14], there is no single unified metric to compare a real-world path which may contain unmapped or dynamic obstacles. While the path lengths of all tasks in this study are comparable and should not contribute to a major difficulty difference. In terms of obstacles, task 1 is very similar to task 2, as they pass through similarly designed corridors and these corridors are kept clutter-free to not hinder personal movement. As such, there may be differences in their decision points or another metric this study has yet to account for. Task 1 is similar to task 3 in the amount of decision points, as well as the locations of these decision points. However, when the participants in task 1 arrive at the decision point in front of room 2017 (see Appendix D), they view a wall in front of them, a path to their right, and a path to the left further ahead. In contrast, the similar decision point of task 3, located in front of G006, features a clear front view, where participants can move forward to check the room number in front of them or turn right and check the rooms down that corridor. This difference may be the cause of the difficulty difference between task 1 and the other tasks.

Unlike the distance results, the completion time performances of each participant do not show a noticeable difference between their average performances. It may be argued that in task 3, the participants who used AR had a better performance on average, and that this difference in performance might be made clearer with more testing. However, it is difficult for this same line of reasoning to be extended to the results of other tasks.

In both time and distance performance metrics, the results of visual AR and audio AR show comparable results while performing better than no AR solutions. This result is in line with previous research, as we expect the performance of audio AR to be similar to or better than visual AR [56; 3; 43], and for audio AR aids to improve the performances of users [37]. As such, we can conclude that RQ1.1 is true in comparison to non-AR users, indicating that audio AR does improve the navigational performance of its users. Such a conclusion cannot be made in comparison to visual AR users. Similarly, RQ1.2 can be refuted when audio AR users are compared against non-AR users, meaning that audio AR does not reduce the performance of its user. Again, such a conclusion cannot be made between audio AR and visual AR.

RQ2 was established to determine how different aspects of AR affect the environmental awareness of people during navigation tasks. To answer this, participants were questioned about the specific and noticeable objects in their environments they were aware of. The results of this task did not show any statistically significant difference between groups. This is likely due to the low number of participants in each group. The results are likely to converge to a more definitive answer with more participants. Additionally, the results show a large confidence window for non-AR users and audio AR users.

One reason for this is due to some participants not paying attention to any objects in their surroundings and focusing on the navigation task. One example of this is P18, who used audio AR and did not notice any objects in their surroundings. At the end of their test, they stated that they did not pay attention to their surroundings. This means that they were not attempting to complete their secondary task, making their result less useful than otherwise. With more participants, this data could be considered to be excluded from the final analysis step to increase the impact of other more reliable participants. In fact with the removal of P18's data, the results of the audio AR group move from being slightly lower than the no AR group with a large confidence interval, to having the same mean and almost the same confidence interval.

Considering the results, it can be shown that people who use audio AR and people who do not use any AR systems are more aware of their surroundings than people who use visual AR systems. If the removal of P18 from the results of this task are accepted, it can be further concluded that people who use audio AR systems are as aware of their surroundings as non-AR users, while visual AR users have a low environmental awareness. Further studies with larger groups may show this relationship more clearly.

These results do not show evidence for RQ 2.1, that audio AR improves the environmental awareness of its users. The results also have shown evidence against RQ2.2, that audio AR does not lower the environmental awareness of its users. On the other hand, there is evidence indicating that visual AR reduces the environmental awareness of its users. The reason for this may be that the visual obstruction of the system is causing attention to be directed away from the environment and onto the virtual display elements. As this distraction does not exist in audio AR systems, the affects of it are also not observed, leading to a level of environmental awareness similar to that of non-AR users.

RQ3 was established to determine how different aspects of AR affect the task load people experience during navigation tasks. This was attempted to be answered by giving users the NASA task load index, which can measure the relevant metrics.

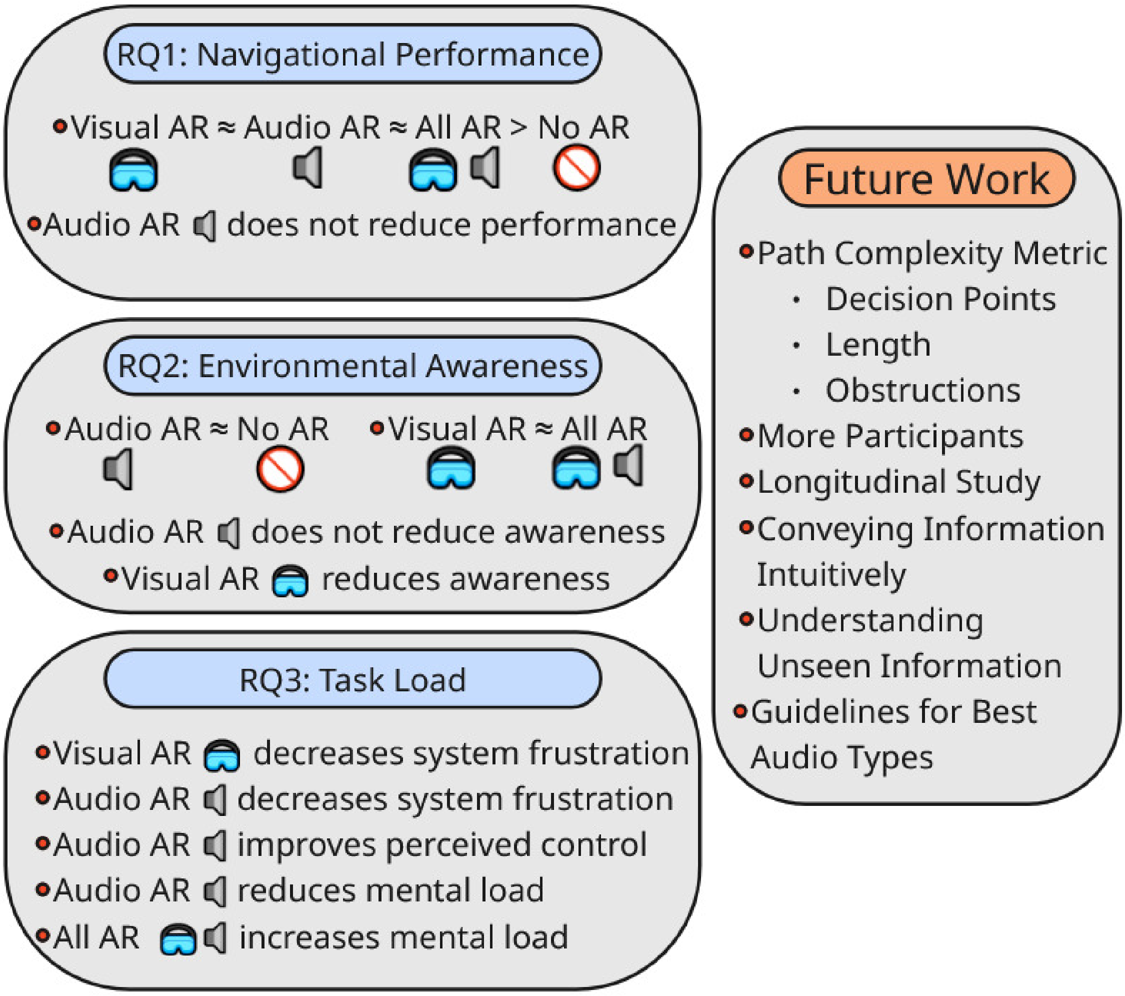
The findings of this index show that audio AR users experienced the lowest mental demand, while non-AR users and visual AR users had similar mental demand. This similarity may be caused many factors, and as such it is difficult to draw direct conclusions from these results in this study. The mental demand results of audio AR users show evidence for RQ3.1, that audio AR improves task load in navigation. It also shows evidence against RQ3.2, that audio AR worsen task load in navigation. These same conclusions cannot be made for visual AR. The mental demand results of visual AR do not show evidence for or against RQ3.1 or RQ3.2. These results are in line with previous research [19], which showed that spatial audio in AR resulted in a lower cognitive load.

In their answers, the participants showed that they felt more in control of their pacing, and the pace of the task in general, when they utilized audio AR. They also showed that when using visual AR and audio AR, they felt that they were able to accomplish their tasks more successfully. Both of these results show evidence for audio AR users having a lower task load. In contrast to these results, the participants reported that when using audio AR, they felt as though they had to spend more effort than the users who did not use any AR and those who used visual AR. This shows that although audio users had to spend more effort to complete their tasks, they still felt in control, completing their tasks at their own pace, and did not experience as much mental demand as other groups. Furthermore, these contrasting findings may be the result of an audio interface being new to the users, which may have caused them to focus more than if they had been experienced with the system. Further longitudinal studies may show a lower effort on the end of audio AR users. Finally, the participants of this study reported that when using visual AR and audio AR, they were not as frustrated as the users who used no AR and those who used all AR. These results provide further evidence for RQ3.1 for both visual AR and audio AR.

While not directly relating to the research questions, the feedback from the participants provides insight into the advantages of audio AR systems. In addition, the comments of P24 indicate that the use of spatialized audio in AR sytems may provide the ability to perceive beyond the immediate area of the user. P24 demonstrated that the were able to perceive and make valid predictions on areas beyond their immediate surroundings, which allowed them to be aware of unseen, far away environments. This is a promising result for the advantages of audio AR in the heightening of the situational awareness of users.

The findings in this study can be summarized as in Figure 5.1. The findings of each research question, as well as the future works that could help improve this area of research are listed individually.

Figure 5.1 Summary of the findings


RQ1: Navigational Performance
•Visual AR ≈ Audio AR ≈ All AR > No AR
•Audio AR does not reduce performance
RQ2: Environmental Awareness
•Audio AR ≈ No AR
•Visual AR ≈ All AR
•Audio AR does not reduce awareness
•Visual AR reduces awareness
RQ3: Task Load
•Visual AR decreases system frustration
•Audio AR decreases system frustration
•Audio AR improves perceived control
•Audio AR reduces mental load
•All AR increases mental load
Future Work
•Path Complexity Metric
• Decision Points
• Length
• Obstructions
•More Participants
•Longitudinal Study
•Conveying Information Intuitively
•Understanding Unseen Information
•Guidelines for Best Audio Types

6. Conclusion

This study investigates the effects of AR systems in the performance and task load of people in navigational tasks within complex, multi-floor structures. More specifically, this study aims to determine how audio AR changes the perception and performance of users in comparison to other forms of navigation in these environments.

To investigate this, a comparative user study was conducted with four experimental conditions: no AR assistance, visual AR assistance, audio AR assistance, and audio and visual AR assistance. The results of these four experimental conditions were then compared based on the four metrics of distance performance, time performance, task load, and environmental awareness.

This study of 28 participants revealed valuable insights on the role of audio in the navigation of complex indoor spaces and how it compares to traditional solutions, which focus on no AR assistance or visual AR assistance. The results of this study have corroborated previous research on audio AR, showing that visual AR users had similar time and distance performances and that audio AR users had the lowest mental load when navigating. The results also provided evidence for audio AR users having a similar environmental awareness to people who do not use any AR assistance, and outperforming the environmental awareness of visual AR users.

The contributions of this study to current literature encompass the use of audio AR in 3D spaces and the impact of audio AR in environmental awareness. As described in Chapter 2, there is currently a lack of studies covering the possible advantages and the effectiveness of audio-based AR interfaces within complex, 3D environments. These environments are generally indoor spaces within buildings. The contribution of this study in this aspect has been the utilization and performance of audio AR in navigational tasks. As a result, evidence has been provided on audio AR performances having consistent performance results with that of 2D navigation.

Furthermore, this study contributes to the literature by providing evidence regarding the environmental awareness of the user when utilizing audio AR. Previous research on audio AR has not analyzed this aspect. As such, there has been uncertainty in the awareness of audio AR users. The results of this study indicate that audio AR is capable of maintaining the environmental awareness of its user to the level of an average person with no AR assistance.

These results indicate that audio AR can support its users to a satisfactory level while not causing major negative side-effects. This means that it can be a valid form of assistance to areas where no forms of AR are utilized traditionally, and that it can replace the use of visual AR in fields where environmental awareness is very important.

For example, first responders can utilize audio AR when navigating dangerous areas. If the structure of the environment is known, audio assistance can provide directives without blocking the visual view of the responder. This can potentially improve their performance. Specifically, police officers may utilize audio over visual AR to ensure that no assailants approach them while they are distracted with the visual interface. Firefighters can utilize this system in low-visibility areas to ensure that they can navigate effectively while their vision is free to spot people in need. Another group of people who can benefit from such a system is the blind and hard of seeing. These users cannot utilize visual AR systems at the level of an average person. Furthermore, not all indoor spaces are built to accommodate these people. As such, audio AR interfaces may be utilized to convey directional information or to inform them of environmental dangers in their immediate vicinity.

7. Limitations and Future Work

This study contains several limitations that may be addressed in future research. The first main limitation of this study setup is the number of participants within each group. Although there were 28 participants, they had to be grouped into four separate experimental conditions. This caused the participant count of each group to be 7 participants. Furthermore, as the study demands physical labor, it is taxing on the researchers to monitor the participants in every test. This causes the study to be conducted a limited number of times every day. In future studies, the number of participants may be increased by conducting the study over a longer period of time. Incentives can also be utilized to draw in more participants.

Another limitation of this study arises from the study design itself. As described in the Methodology, there had to be made some assumptions in what audio type should be selected. In the literature, there is no specific study on what guidelines to follow for the selection of audio-types on a per-task basis. Although there is research in the characteristics the audio should have, other researchers have to semi-arbitrarily choose certain audio types in their design. As such, future research in the field would benefit greatly from a more concrete guideline to follow in how they should design their audio tasks. This can help in choosing what audio types are more appropriate over others for a given task, specific conditions, or a specific group of people.

In the analysis of this study, some results indicated that one task out of the four given to a participant might have been easier to complete. This conclusion cannot be made concretely as there is no numerical method of evaluating a given path on its difficulty level. Due to this, this study cannot guarantee that all task were similar. In future works, a measure can be developed which can evaluate the difficulty level of a given path. Such a measure can help with better understanding the phenomena encountered in this study or similar phenomena in other research. Such a measure can also help with choosing different paths in navigation tasks, where each of them can be more reliably compared against each other.

This study had to make assumptions on the method of hearing its participants utilized. As described in the background of the study, generic HRTF devices can be used by anyone and still convey useful information to its user. However, specialized HRTF devices, which conform to the specific head shape and ear structure of the user, can still boost the performance of its user. Such a change may exasperate the results found in this study. Specialized HRTF devices don't have to be hardware, as software simulations of the user's head can give satisfying results as well. To do this, the extensive work of [17] can be utilized. This dataset contains many variations of head shapes and their acoustic characteristics. While it may not be able to directly match a user's head, it can allow a study to use better approximations than that of a generic HRTF.

Finally, further research may also focus on the creation of audio guidelines for researchers to follow. In this study, the audio sound of a radar ping was picked due to its association with a search task. However, this sound can be too loud or ambiguous for some users. This is also a somewhat arbitrary choice, as there is no hard proof that this is the most suitable sound choice. A guideline that can be referenced by researchers when deciding what sound to place into their environment can alleviate this ambiguity. This may be done through consulting with audio designers, sound engineers, or game designers who regularly have to make decisions on what a user should hear in a certain situation.

BIBLIOGRAPHY

[1] Zeynep Abes, Nathan Fairchild, Spencer Lin, Michael Wahba, Katrina Xiao, and Scott S. Fisher. The immersive archive: Archival strategies for the sensorama. In *2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, pages 307–312. IEEE, January 2025. doi: 10.1109/aixvr63409.2025.00059. URL http://dx.doi.org/10.1109/AIxVR63409.2025.00059.

[2] Adam Glass and Melanie Kneitmix. Soundscape, May 2025. URL https://github.com/microsoft/soundscape.

[3] Barde Amit, Ward Matt, Lindeman Robert, and Billinghurst Mark. The use of spatialised auditory and visual cues for target acqusition in a search task. *Journal of the Audio Engineering Society*, (2-5), August 2020.

[4] Izar Azpiroz, Igor Garcia Olaizola, Xabier Oregui, Anaida Fernandez Garcia, Veronica Ruiz, Blanca Larraga-Garcia, and ÃĄlvaro Gutierrez. White paper on adaptive situational awareness enhancing augmented reality interface design on first responders in rescue tasks. *Applied Sciences*, 14(18), 2024. ISSN 2076-3417. doi: 10.3390/app14188282. URL https://www.mdpi.com/2076-3417/14/18/8282.

[5] Ronald T. Azuma. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, August 1997. doi: 10.1162/pres.1997.6.4.355. URL https://doi.org/10.1162/pres.1997.6.4.355.

[6] Stephen Barrass and Gregory Kramer. Using sonification. *Multimedia Systems*, 7(1):23–31, January 1999. ISSN 1432-1882. doi: 10.1007/s005300050108. URL https://doi.org/10.1007/s005300050108.

[7] Durand R. Begault. *3-D sound for virtual reality and multimedia*. Academic Press Professional, Inc., San Diego, CA, USA, August 1994. ISBN 0-12-084735-3. URL https://dl.acm.org/doi/book/10.5555/184407.

[8] Jens Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, October 1996. ISBN 978-0-262-26868-4. doi: 10.7551/mitpress/6391.001.0001. URL https://doi.org/10.7551/mitpress/6391.001.0001.

[9] Robert A. Butler. The relative influence of pitch and timbre on the apparent location of sound in the median sagittal plane. *Perception & Psychophysics*, 14(2):255–258, June 1973. ISSN 1532-5962. doi: 10.3758/BF03212386. URL https://doi.org/10.3758/BF03212386.

[10] Todd Carpenter. Pixabay - royalty free image and audio, November 2010. URL https://pixabay.com/.

[11] T.P. Caudell and D.W. Mizell. Augmented Reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, volume ii, pages 659–669 vol.2, Kauai, HI, USA, January 1992. IEEE. ISBN 0-8186-2420-5. doi: 10.1109/HICSS.1992.183317. URL https://ieeexplore.ieee.org/document/183317.

[12] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch. Comparing the Effect of Audio and Visual Notifications on Workspace Awareness Using Head-Mounted Displays for Remote Collaboration in Augmented Reality. *Augmented Human Research*, 1(1):1, October 2016. ISSN 2365-4325. doi: 10.1007/s41133-016-0003-x. URL https://doi.org/10.1007/s41133-016-0003-x.

[13] Gregory D. Clemenson, Antonella Maselli, Alexander J. Fiannaca, Amos Miller, and Mar Gonzalez-Franco. Rethinking GPS navigation: Creating cognitive maps through auditory clues. *Sci Rep*, 11(1):7764, April 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-87148-4. URL https://www.nature.com/articles/s41598-021-87148-4. Nature Publishing Group.

[14] Hannah Jacqueline A. Dasal, Ma Ericka G. Gutierrez, Ma Krizel Anne V. Zulueta, Leisyl M. Mahusay, Elsa S. Pascual, Joshua James D. Magora, Jamillah S. Guialil, and Jonathan C. Morano. Enhancement of Harris Hawks optimization applied in path planning for an indoor navigation mobile application. *WJARR*, 22(2):681–700, 2024. ISSN 2581-9615. doi: 10.30574/wjarr.2024.22.2.1424. URL https://wjarr.com/content/enhancement-harris-hawks-optimization-applied-path-planning-indoor-navigation-mobile.

[15] Joseph L. Gabbard, Divya Gupta Mehra, and J. Edward Swan. Effects of AR display context switching and focal distance switching on human performance. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2228–2241, 2019. doi: 10.1109/TVCG.2018.2832633.

[16] Zihan Gao, Huiqiang Wang, Guangsheng Feng, and Hongwu Lv. Exploring sonification mapping strategies for spatial auditory guidance in immersive virtual environments. *ACM Trans. Appl. Percept.*, 19(3), September 2022. ISSN 1544-3558. doi: 10.1145/3528171. URL https://doi.org/10.1145/3528171.

[17] Bill GARDNER and Keith Martin. HRTF Measurements of a KEMAR Dummy-Head Microphone. *MIT Media Lab. Perceptual Computing-Technical Report*, 280:1–7, May 1994. URL http://www.linux.bucknell.edu/~kozick/elec32007/hrtfdoc.pdf.

[18] Mehmet Göktürk and Nihat Erim İnceoğlu. Real Time Sensory Substitution for the Blind. In *HCI International 2007*, volume 4550-4566, Beijing, China, July 2007. Springer. URL https://www.academia.edu/9309745/REAL_TIME_SENSORY_SUBSTITUTION_FOR_THE_BLIND.

[19] Jeremy Raboff Gordon, Alexander J. Fiannaca, Melanie Kneisel, Edward Cutrell, Amos Miller, and Mar Gonzalez-Franco. Hearing the way forward: Exploring ambient navigational awareness with reduced cognitive load through spatial audio-AR. In *Extended Abstracts of the 2023 CHI Conference on*

Human Factors in Computing Systems, CHI EA '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394222. doi: 10.1145/3544549.3585800. URL https://doi.org/10.1145/3544549.3585800.

[20] GraphPad Software Inc. Prism 10.4.2 (633), March 2025. URL https://www.graphpad.com.

[21] Hans Fugal, Varun Nair, and Paul O'Shannessy. Facebook 360 Spatial Workstation, May 2025. URL https://github.com/facebookarchive/facebook-360-spatial-workstation.

[22] Sandra G. Hart. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50 (9):904–908, October 2006. doi: 10.1177/154193120605000909. URL https://doi.org/10.1177/154193120605000909.

[23] Morton L. Heilig. Sensorama simulator. U.S. Patent No. US3050870A. New York. U.S. Patent and Trademark Office, August 1962. URL https://patents.google.com/patent/US3050870A/en.

[24] Florian Heller, Aaron Krämer, and Jan Borchers. Simplifying orientation measurement for mobile audio augmented reality applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 615–624, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450324731. doi: 10.1145/2556288.2557021. URL https://doi.org/10.1145/2556288.2557021.

[25] Simon Holland, David R. Morse, and Henrik Gedenryd. AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. *Personal and Ubiquitous Computing*, 6(4):253–259, September 2002. ISSN 1617-4909. doi: 10.1007/s007790200025. URL https://doi.org/10.1007/s007790200025.

[26] hth2000 (Freesound). Sonar | Royalty-free Music, August 2022. URL https://pixabay.com/sound-effects/sonar-89448/.

[27] Nicholas Huang and Mounya Elhilali. Auditory salience using natural soundscapes. *J Acoust Soc Am*, 141(3):2163, March 2017. ISSN 1520-8524 0001-4966. doi: 10.1121/1.4979055.

[28] Katelynn Kapalo, Patricia Bockelman, and Joseph LaViola. Sizing up emerging technology for firefighting: Augmented reality for incident assessment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62: 1464–1468, 09 2018. doi: 10.1177/1541931218621332.

[29] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, pages 85–94, San Francisco, CA, USA, October 1999. IEEE. doi: 10.1109/IWAR.1999.803809. URL https://ieeexplore.ieee.org/document/803809.

[30] You Jin Kim, Radha Kumaran, Ehsan Sayyad, Anne Milner, Tom Bullock, Barry Giesbrecht, and Tobias Hollerer. Investigating search among physical and virtual objects under different lighting conditions. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3788–3798, 2022. doi: 10.1109/TVCG.2022.3203093.

[31] Sandeep Reddy Kothinti, Nicholas Huang, and Mounya Elhilali. Auditory salience using natural scenes: An online study. *J Acoust Soc Am*, 150(4):2952, October 2021. ISSN 1520-8524 0001-4966. doi: 10.1121/10.0006750. United States.

[32] J. Laird Evans and A. Stevens. Measures of graphical complexity for navigation and route guidance displays. *Displays*, 17(2):89–93, 1997. ISSN 0141-9382. doi: https://doi.org/10.1016/S0141-9382(96)01025-6. URL https://www.sciencedirect.com/science/article/pii/S0141938296010256.

[33] Philip Lamb, Guillaume Martres, and Raphael Druon. artoolkitX, May 2025. URL https://github.com/artoolkitx/artoolkitx.

[34] Hyunkook Lee. Phantom image elevation explained. *AES*, (141):9664, September 2016. URL http://eprints.hud.ac.uk/id/eprint/29581/.

[35] Tiffany Liu, Javier Hernandez, Mar Gonzalez-Franco, Antonella Maselli, Melanie Kneisel, Adam Glass, Jarnail Chudge, and Amos Miller. Characterizing and predicting engagement of blind and low-vision people with an audio-based navigation app. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391566. doi: 10.1145/3491101.3519862. URL https://doi.org/10.1145/3491101.3519862.

[36] Marcin Gorzel, Eric Mauskopf, Julius Kammerl, and Clayton Wilkinson. Resonance Audio, November 2023. URL https://github.com/resonance-audio/resonance-audio.

[37] Richard L. McKinley and Mark A. Ericson. Flight demonstration of a 3-D auditory display. In *Binaural and spatial hearing in real and virtual environments.*, pages 683–699. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1997. ISBN 0-8058-1654-2 (Hardcover).

[38] Ahmadreza Meysami, Sousso Kelouwani, Jean-Christophe Cuilliere, Vincent Francois, Ali Amamou, and Bilel Allani. An efficient indoor large map global path planning for robot navigation. *Expert Systems with Applications*, 248:123388, 2024. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2024.123388. URL https://www.sciencedirect.com/science/article/pii/S0957417424002537.

[39] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Trans. Information Systems*, vol. E77-D, no. 12:1321–1329, 12 1994.

[40] Maksims Mironovs and Hyunkook Lee. On the accuracy and consistency of sound localisation at various azimuth and elevation angles. *AES*, 144(9952), May 2018. URL https://aes2.org/publications/elibrary-page/?id=19469.

[41] Michael Nees and Bruce Walker. Theory of Sonification. In *Principles of Sonification: An Introduction to Auditory Display*, volume 1, pages 9–39. Logos Publishing House, Berlin, January 2012. ISBN 978-3-8325-2819-5.

[42] Sebeom Park, Shokhrukh Bokijonov, and Yosoon Choi. Review of Microsoft HoloLens Applications over the Past Five Years. *Applied Sciences*, 11:7259, August 2021. doi: 10.3390/app11167259.

[43] Simon P. A. Parker, Sean E. Smith, Karen L. Stephan, Russell L. Martin, and Ken I. McAnally and. Effects of supplementing head-down displays with 3-D audio during visual target acquisition. *The International Journal of Aviation Psychology*, 14(3):277–295, 2004. doi: 10.1207/s15327108ijap1403_4. URL https://doi.org/10.1207/s15327108ijap1403_4.

[44] Anna Preis, Jedrzej Kocinski, Honorata Hafke-Dys, and Malgorzata Wrzosek. Audio-visual interactions in environment assessment. *Science of The Total Environment*, 523:191–200, 2015. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2015.03.128. URL https://www.sciencedirect.com/science/article/pii/S004896971500412X.

[45] Vani G. Rajendran and Hannes Gamper. Spectral manipulation improves elevation perception with non-individualized head-related transfer functions. *J Acoust Soc Am*, 145(3):EL222, March 2019. ISSN 1520-8524 0001-4966. doi: 10.1121/1.5093641.

[46] Samuel Sandberg, Calle Hakansson, Niklas Elmqvist, Philippas Tsigas, and Fang Chen. Using 3D audio guidance to locate indoor static objects. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(16):1581–1584, 2006. doi: 10.1177/154193120605001614. URL https://doi.org/10.1177/154193120605001614.

[47] Joram Schito and Sara Irina Fabrikant. Exploring maps by sounds: Using parameter mapping sonification to make digital elevation models audible. *International Journal of Geographical Information Science*, 32(5):874–906, 2018. doi: 10.1080/13658816.2017.1420192. URL https://doi.org/10.1080/13658816.2017.1420192.

[48] Mel Slater and Martin Usoh. Representations Systems, Perceptual Position, and Presence in Immersive Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 2(3):221–233, August 1993. doi: 10.1162/pres.1993.2.3.221. URL https://doi.org/10.1162/pres.1993.2.3.221.

[49] V. Sundareswaran, K. Wang, S. Chen, R. Behringer, J. McGee, C. Tam, and P. Zahorik. 3D audio augmented reality: Implementation and experiments. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pages 296–297, Oct 2003. doi: 10.1109/ISMAR.2003.1240728.

[50] Rudolf Susnik, Jaka Sodnik, and Saso Tomazic. Coding of elevation in acoustic image of space. In *Acoustics: Proceedings of the Annual Conference of the Australian Acoustical Society 2005*, pages 145–50, Busselton, Western Australia, November 2005. Citeseer.

[51] The IEEE and The Open Group. The Open Group Base Specifications Issue 8, IEEE Std 1003.1-2024. https://pubs.opengroup.org/onlinepubs/9799919799/, 2024. Chapter 4, Section 19.

[52] Thomas Hermann, Andy Hunt, and John G. Neuhoff. *The Sonification Handbook*. Logos Publishing House, Berlin, 1 edition, January 2011. ISBN 978-3-8325-2819-5. URL https://sonification.de/handbook/.

[53] Rick Van Krevelen and Ronald Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1–20, June 2010. doi: 10.20870/IJVR.2010.9.2.2767.

[54] Yolanda Vazquez-Alvarez, Ian Oakley, and Stephen A. Brewster. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous Computing*, 16(8):987–999, December 2012. ISSN 1617-4917. doi: 10.1007/s00779-011-0459-0. URL https://doi.org/10.1007/s00779-011-0459-0.

[55] vradmin. VPL Research Jaron Lanier, September 2013. URL https://www.vrs.org.uk/virtual-reality-profiles/vpl-research.html.

[56] Bruce N. Walker and Jeffrey Lindsay. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Hum Factors*, 48(2):265–278, 2006. ISSN 0018-7208. doi: 10.1518/001872006777724507.

[57] Stanley Grauman Weinbaum. *Pygmalion's Spectacles*. Kessinger Publishing, June 1935. ISBN 978-1-4191-4352-6. URL http://archive.org/details/pygmalionsspecta22893gut.

[58] John Werner. Catchup With Ivan Sutherland - Inventor Of The First AR Headset, February 2024. URL https://www.forbes.com/sites/johnwerner/2024/02/23/catchup-with-ivan-sutherlandinventor-of-the-first-ar-headset/. Section: AI.

[59] Jing Yang, Amit Barde, and Mark Billinghurst. Audio Augmented Reality: A Systematic Review of Technologies, Applications, and Future Research Directions. *AES*, 70(10):788–809, October 2022. doi: 10.17743/jaes.2022.0048. URL https://secure.aes.org/forum/pubs/journal/?elib=22008.

[60] Kexin Zhang, Brianna R Cochran, Ruijia Chen, Lance Hartung, Bryce Sprecher, Ross Tredinnick, Kevin Ponto, Suman Banerjee, and Yuhang Zhao. Exploring the design space of optical see-through AR head-mounted displays to support first responders in the field. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642195. URL https://doi.org/10.1145/3613904.3642195.

[61] Yuhang Zhao, Elizabeth Kupferstein, Hathaitorn Rojnirun, Leah Findlater, and Shiri Azenkot. The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–14, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. doi: 10.1145/3313831.3376516. URL https://doi.org/10.1145/3313831.3376516.

[62] Jingyi Zhou, Jie Shen, Cheng Fu, Robert Weibel, and Zhiyong Zhou. Quantifying indoor navigation map information considering the dynamic map elements for scale adaptation. *International Journal of Applied Earth Observation and Geoinformation*, 136:104323, 2025. ISSN 1569-8432. doi: https://doi.org/10.1016/j.jag.2024.104323. URL https://www.sciencedirect.com/science/article/pii/S1569843224006812.

[63] Jingyi Zhou, Haoyu Yang, Jie Shen, and Litao Zhu and. Indoor navigation map design based on spatial complexity. *Cartography and Geographic Information Science*, 52(1):69–81, 2025. doi: 10.1080/15230406.2024.2339296. URL https://doi.org/10.1080/15230406.2024.2339296.

APPENDIX A

NASA Task Load Index

How mentally demanding was the task?	- - -	- -	-	-/+	+	+ +	+ + +	
Very Low	□	□	□	□	□	□	□	Very High

How physically demanding was the task?	- - -	- -	-	-/+	+	+ +	+ + +	
Very Low	□	□	□	□	□	□	□	Very High

How hurried or rushed was the pace of the task?	- - -	- -	-	-/+	+	+ +	+ + +	
Very Low	□	□	□	□	□	□	□	Very High

How successful were you in accomplishing what you were asked to do?	- - -	- -	-	-/+	+	+ +	+ + +	
Very Low	□	□	□	□	□	□	□	Very High

How hard did you have to work to accomplish your level of performance	- - -	- -	-	-/+	+	+ +	+ + +	
Very Low	□	□	□	□	□	□	□	Very High

How insecure, discouraged, irritated, stressed, and annoyed were you?	- - -	- -	-	-/+	+	+ +	+ + +	
Very Low	□	□	□	□	□	□	□	Very High

APPENDIX B

Environmental Awareness Test Objects

The items that were actually in the physical world are marked with *.

*A bright yellow trash box.

*An orange large backpack.

*A black coat hanger.

*A white stand.

*A cardboard box.

A large 3-liter bottle of black cola soft drink.

A large black suitcase.

A 5-liter water bottle.

A dark green mesh can.

A large 3-liter bottle of orange carbonated soft drink.

A small fridge.

Basketball.

A cart with books.

A light stand.

A sewing machine.

APPENDIX C

Post-Test Questionnaire

What is your experiment ID?

____

What is your age?

____

What is your gender?	Male	Female	Other
	□	□	□

What's your department? (eg. Faculty of Engineering and natural sciences - FENS)

____

On a scale from 1 to 5, where 1 is not experienced at all and 5 is very experienced, what is your level of experience with Virtual Reality or Augmented Reality devices? (eg. Quest, Vive, Hololens)	1	2	3	4	5	
No experience	□	□	□	□	□	Very experienced

On a scale from 1 to 5, where 1 means are not familiar at all and 5 means you are very familiar, how familiar are you with the building you are navigating in this study?	1	2	3	4	5	
No experience	□	□	□	□	□	Very experienced

On a scale from 1 to 5, where 1 is very poor and 5 is very skilled, how would you rate your navigational skills?	1	2	3	4	5	
No experience	□	□	□	□	□	Very experienced

Do you have any notable disabilities, especially in sight or hearing?

____

How often do you play video games?

Not at all □
Once every month (rarely) □
Once every two weeks (semi-rarely) □
Once every week (occasionally) □
up to three times a week (semi-often) □
more than three times a week (often) □

In the previous question, if you stated you play games, please specify what kinds of video games you play. (eg. first person shooters, sports, real time strategy)

____

Do you do any sports which may require navigational or tracking skills? (eg. hiking, orienteering, scouting)

____

Do you have any notes on how you felt while completing your tasks?

____

Do you have any extra notes?

____

APPENDIX D

Task Maps

In the following figures, the red triangle represents the starting position of each participant, the red square represents the target location, the red line represents the ideal path from start to finish, and the green circles represent the items on the paths placed for the environmental awareness test. Along with these symbols, the floor level each image displays is written in text in the middle of each image. Each figure represents a different targets navigation task.

Figure D.1 Task 1 Map

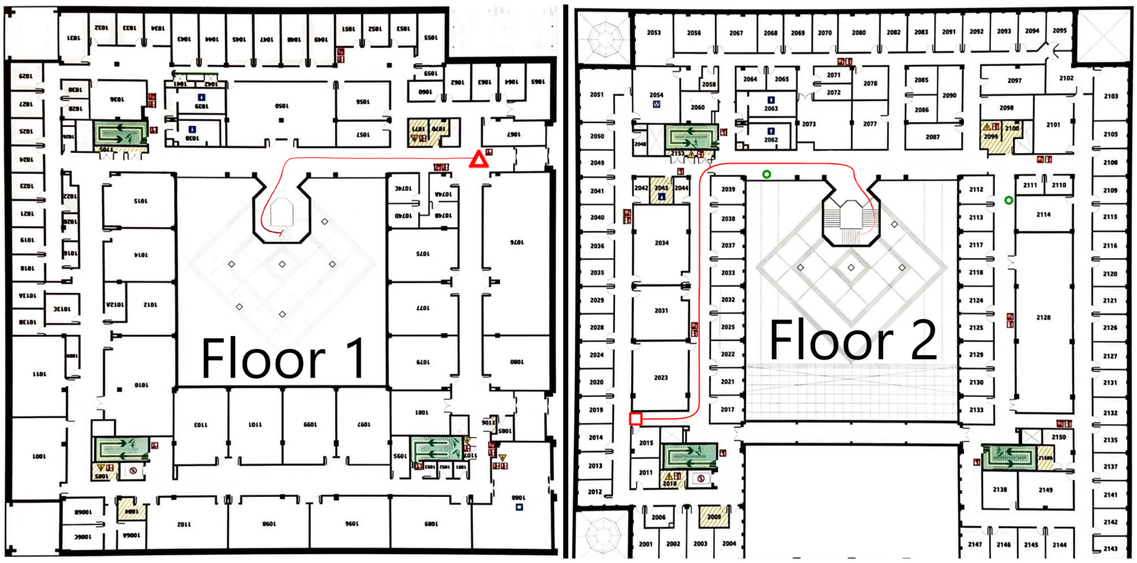


Figure D.2 Task 2 Map

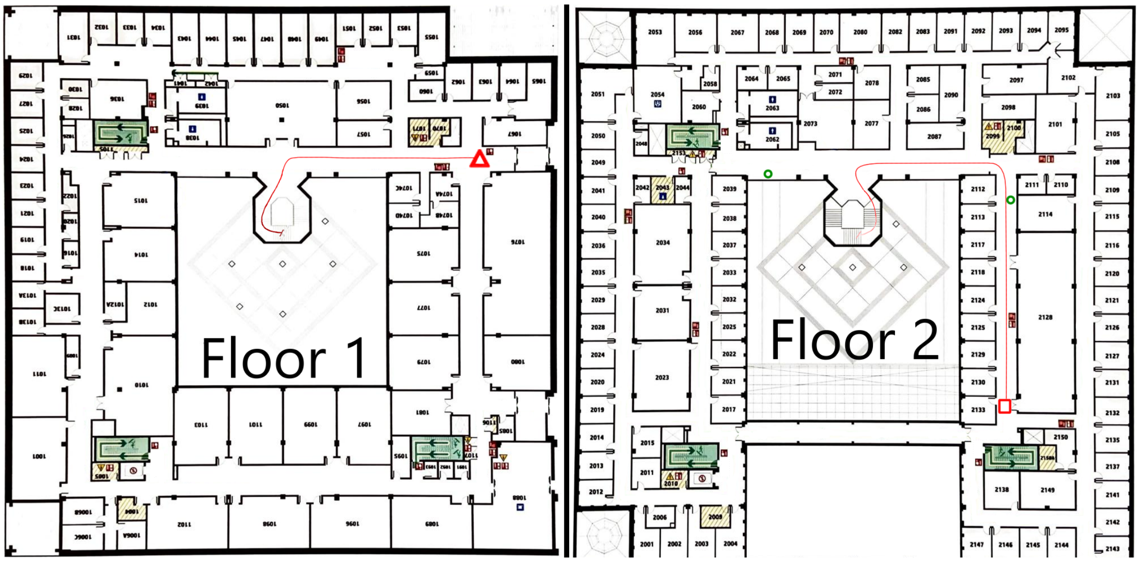


Figure D.3 Task 3 Map

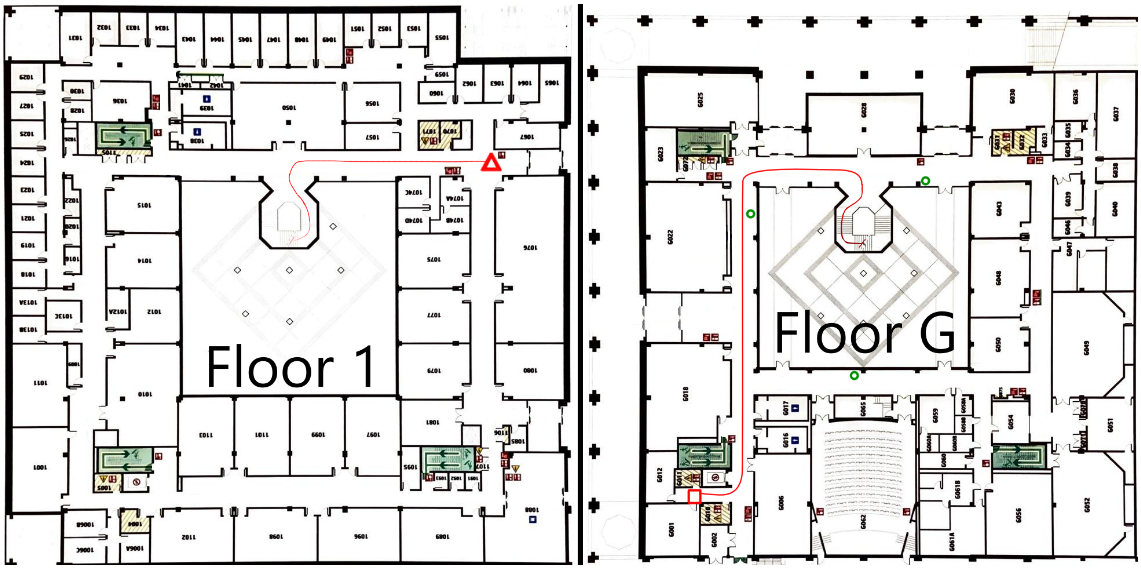
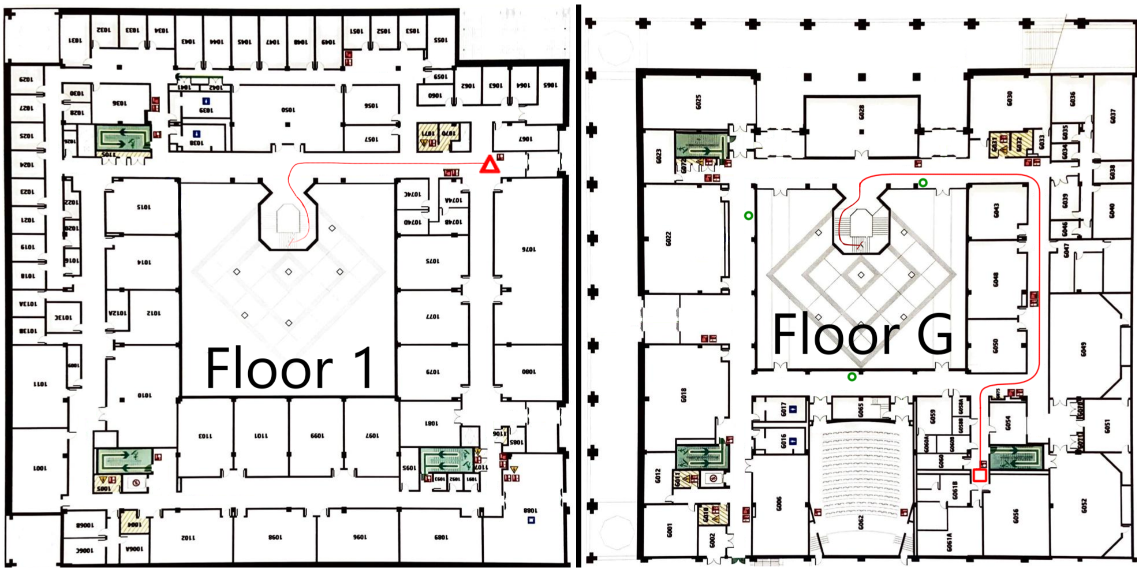


Figure D.4 Task 4 Map



APPENDIX E

Blender Code

```
import bpy
import csv

# Convert to absolute path
csv_filepath = bpy.path.abspath("//ID/ID_TestNum.csv")

# Column names
# y-z are swapped since blender uses z as height
X_COLUMN = "position x"
Z_COLUMN = "position y"
Y_COLUMN = "position z"

# Lists to store the point data
verts = []
edges = []

# Read CSV, extract point data
with open(csv_filepath, newline=’’) as csvfile:
    reader = csv.DictReader(csvfile)  # Read as dictionary
    prev_index = None

    for index, row in enumerate(reader):
        try:
            x, y, z = float(row[X_COLUMN]), float(row[Y_COLUMN]),
                                              float(row[Z_COLUMN])
            verts.append((x, y, z))  # Store vertex

            # Create edge from previous vertex to current
            if prev_index is not None:
                edges.append((prev_index, index))

            prev_index = index  # Update the last index
        except (ValueError, KeyError) as e:
            print(f"Skipping row {index} due to error: {e}")

# Create the mesh and object in Blender
mesh = bpy.data.meshes.new("PointPath")
mesh.from_pydata(verts, edges, [])  # Create mesh with vertices &
                                  edges
mesh.update()
```

```
# Add custom attribute to store vertex index
attr = mesh.attributes.new(name="vertex_index", type=’INT’, domain=
                                      ’POINT’)
for i in range(len(verts)):
    attr.data[i].value = i  # Store index in attribute

obj = bpy.data.objects.new("PointPath", mesh)
bpy.context.collection.objects.link(obj)  # Link object to the
                                      scene

# Create text objects for vertex labels
def create_text_label(text, location):
    txt_data = bpy.data.curves.new(name=f"Label_{text}", type=’FONT
                                          ’)
    txt_obj = bpy.data.objects.new(f"Label_{text}", txt_data)
    txt_data.body = text
    txt_obj.location = location
    txt_obj.scale = (0.005, 0.005, 0.005)  # Adjust text size
    bpy.context.collection.objects.link(txt_obj)

# Add labels above each vertex
for i, (x, y, z) in enumerate(verts):
    create_text_label(str(i), (x, y, z + 0.0001))
```

APPENDIX F

Performance Analysis Code

```python
import pandas as pd
import numpy as np

# Load CSV file
df = pd.read_csv("./ID/ID_TestNum.csv")

# Identify valid rows
df["group"] = (df["test Started"] == 0).cumsum()

# Filter out invalid rows
valid_groups = df[df["test Started"] == 1].groupby("group")

# Store results
segments = [] # stores the individual valid data groups
results = [] # stores (elapsed_time, total_distance) for segments

for _, group in valid_groups:
    # Convert to numpy array
    data = group.drop(columns=["group"]).to_numpy()

    # Calculate elapsed time
    elapsed_time = data[-1, 0] - data[0, 0]  # col 0 is timestamp

    # Calculate total distance traveled (Euclidean sum)
    coords = data[:, 2:5]  # x, y, z are in columns 1, 2, 3
    distances = np.sqrt(np.sum(np.diff(coords, axis=0) ** 2, axis=1
                                     ))
    total_distance = np.sum(distances)

    # Store segment and results
    segments.append(data)
    results.append((elapsed_time, total_distance))

totTime = 0
totDist = 0

for i, (elapsed, distance) in enumerate(results):
    totTime += elapsed
    totDist += distance
print(f"Elapsed Time = {totTime} s, Total Distance = {totDist} m")
```

APPENDIX G

Ethics Committee Approval Form

The original ethics committee approval is attached below.

: 18.04.2025-50648

Orhanlı
81474 Tuzla / İstanbul –Turkey
Telefon / Phone:+90 (0216) 483 9000
Faks / Facsimile:+90 (0216) 483 9005

İletişim Merkezi/ Communication Center:
Bankalar Caddesi 2
80020 Karaköy / İstanbul – Turkey
Telefon / Phone:+90 (0212) 292 4939
Faks / Facsimile:+90 (0212) 252 3293

www.sabanciuniv.edu

TARİH : 18.04.2025

YER : Sabancı Üniversitesi, Orta Mah. Üniversite Cad. No:27, Tuzla 34956 Istanbul

KATILIMCILAR : Canan Atılgan, Cemal Yılmaz,Umut Şahin,Zafer Gedik

SABANCI ÜNİVERSİTESİ ARAŞTIRMA ETİK KURULU (AEK) ONAY FORMU

PROJENİN /ARAŞTIRMANIN ADI	Karmaşık Çok Katlı Ortamlarda Ses Öncelikli Karma Gerçeklik Kullanılarak Navigasyon Performansı
PROJENİN /ARAŞTIRMANIN YÜRÜTÜCÜSÜ, İLETİŞİM BİLGİLERİ VE EKİBİ	Selim Balcısoy: balcisoy@sabanciuniv.edu +905336640363 Bilgehan Çağıltay: bcagiltay@sabanciuniv.edu +905446503594
PROJENİN /ARAŞTIRMANIN AEK'YA BAŞVURMA NEDENİ	Proje kapsamında katılımcıların kişisel bilgilerinin kaydedilmesi ve işlenmesi
PROJE/ ARAŞTIRMA BİR KURULUŞ TARAFINDAN DESTEKLENİYOR MU?	**HAYIR**
PROJENİN /ARAŞTIRMANIN BAŞLANGIÇ TARİHİ	14/04/2025

PROJENİN /ARAŞTIRMANIN AMACI	Bu projede, çok katlı binalarda kullanıldığında sesli artırılmış gerçekliğin sayesinde kullanıcıların çevresel farkındalığını artırma yöntemlerini araştırıyoruz. Sesli simgeler ile yönlendirilirken kullanıcı davranışını analiz edeceğiz, bu sistemin temel avantajlarını ve dezavantajlarını belirleyeceğiz ve bunun sektördeki mevcut teknolojilere ve uygulamalara nasıl entegre edilebileceğini göstereceğiz.
PROJENİN /ARAŞTIRMANIN ETİK İLE İLGİLİ GEREKÇESİ	Proje, 80 kişilik bir katılımcı grubunun kişisel bilgilerini gerektirecek. Bu bilgilerin gizliliği ve etik açısından uygun olması büyük önem taşımaktadır. Bu sebeple etik kurul izni ile bu bilgilerin gizliliğinin korunacağını ve yalnızca araştırma amaçları için kullanılacağının tasdik edilmesi gerekmektedir.
PROJENİN /ARAŞTIRMANIN YÖNTEMİ	Sesli ve görsel arttırılmış gerçeklik kullanıcıların ortam ve yapılması gereken görev hakkında daha fazla bilgi sahibi olmasını sağlar. Testin tamamlanması için katılımcılar sistem testi ve anket doldurma aşamalarını tamamlamalıdır. Katılımcılar Sabancı Üniversitesi lisans ve lisansüstü öğrencilerinden oluşacak ve testler İngilizce olarak gerçekleştirilecektir. Sistem testleri tamamen şahsen, Sabancı Üniversitesi kampüsü, Sanat ve Sosyal Bilimler Fakültesi binasında gerçekleştirilecektir. Katılımcılara kulaktan kulağa haber dağılımı ve üniversite kampüsü içindeki bağlantılar aracılığıyla ulaşılacaktır. Katılımcıların hepsi 18 yaşından büyük olacaktır. Sistem testi ve anketin süresi toplamda 30-40 dakika sürecektir. Sistem test süreci sırasında, kullanıcılar konumlarının takip edilmesi amacıyla karma gerçeklik başlığı takacaklardır. Katılımcılardan fakülte binası içindeki konumlara fiziksel olarak gitmeleri istenecektir. Gezinmeleri sırasında, çevresel farkındalıklarını ölçmek için çevrelerindeki belirli ögelerle ilgili sorular yanıtlamak üzere periyodik olarak duracaklardır. Başlık şeffaf ekran kullandığı ve katılımcıların görsel alanını engellemediği için dış ortamlarını görememe riskiyle karşı karşıya kalmayacaklardır.

ETİK İLE İLGİLİ KULLANILACAK BİYOLOJİK, PSİKOLOJİK VE TEKNİK VB TÜM YÖNTEMLER	Projede kullanılacak veri katılımcıların kişisel verilerini ve test sırasındaki hareketlerini içermektedir. Kayıtlarının kaydedilmesini kabul eden kullanıcılar çalışmada cinsiyet, yaş, fakülte ve video oyunları oynuyorlarsa ne oynadıklarını bildirmeleri gerekmektedir. Deney sırasında katılımcıların 3 boyutlu uzayda kafalarının hangi koordinatta bulunduğu (x, y, z koordinatlarında) kaydedilecektir. Katılımcılar sadece kendilerine deney başında verilen kimlik numaraları ile kaydedilecektir. Proje boyunca veri sadece Selim Balcısoy ve Bilgehan Çağıltay'ın erişiminde olacaktır. Toplanılan veriler yerel olarak depolanacak ve internet üzerinden erişilemeyecektir. Depolama aygıtı şifre korumalı olacaktır. Veriler gelecekteki analizler için süresiz olarak saklanacaktır.
ETİK İLE İLGİLİ KULLANILACAK PROSEDÜR VE İLGİLİ RİSKLER YA DA TEHDİTLER	Katılımcının bilgilerini açık edebilecek herhangi bir öznitelik (TCKN, adres, telefon numarası vb.) bulunmamaktadır. Her bir satırdaki öznitelikler (cinsiyet, yaş, fakülte) belirli bir katılımcıya aittir. Bu nedenle kullanılan veri, katılımcıların gizliliğine herhangi bir tehdit oluşturmamaktadır.
RİSKLER YA DA TEHDİTLERİ ENGELLEYECEK ÖNLEMLER NELERDİR?	Kullanılan veri, kullanıcının gizliliğine herhangi bir tehdit oluşturmamaktadır.
PROJENİN /ARAŞTIRMANIN İÇERİĞİNDE HERHANGİ BİR ÖDÜL YA DA ÜCRET UYGULAMASI OLUP OLMADIĞINI AÇIKLAYINIZ	Katılımcılara herhangi bir katılım teşvik edebilecek ödül ya da ücret uygulaması olmayacaktır.
VERİ SAĞLANACAK KİŞİLERDEN/EVEBEYNLERDEN BU ÇALIŞMA İÇİN İZİN FORMU ALINDI MI?	Hayır.
BU ÇALIŞMANIN YAPILACAĞI BAŞKA KURUM VARSA O KURUMDAN ONAY ALINDI MI?	Hayır, başka bir kurum yoktur.

Sabancı Üniversitesi MDBF öğretim üyelerinden Prof. Selim Bacısoy'un Karmaşık Çok Katlı Ortamlarda Ses Öncelikli Karma Gerçeklik Kullanılarak Navigasyon Performansı adlı projesi/araştırması AEK tarafından değerlendirilmiştir.

Proje etik açısından uygun bulunmuştur. ■

Projenin etik açısından geliştirilmesi gerekmektedir. □

Proje etik açısından uygun bulunmamıştır. □

KONTROL LİSTESİ	**EVET**	**HAYIR**
AEK'YA BAŞVURMA NEDENİ AYRINTILI BİR ŞEKİLDE BELİRTİLDİ Mİ?	x	
YAPILACAK ARAŞTIRMA KONUSU İLE İLGİLİ BİLGİ VERİLDİ Mİ?	x	
VERİLERİN NE ŞEKİLDE ELDE EDİLECEĞİ ANLATILDI MI?	x	
VERİLERİN KİMLERDEN/NEREDEN VE NASIL ALINACAĞI AÇIKLANDI MI? (KİŞİLER SÖZ KONUSU İSE GÖNÜLLÜLÜK YA DA HERHANGİ BİR ÖDÜL SİSTEMİ OLUP OLMADIĞI VB)	x	
ELDE EDİLECEK VERİLERİN NE KADAR SÜRE İLE ARŞİVLENECEĞİNE YÖNELİK BİLGİ VERİLDİ Mİ?	x	
VERİLERİN KİMLERLE PAYLAŞILACAĞI NET BİR ŞEKİLDE AÇIKLANDI MI?	x	