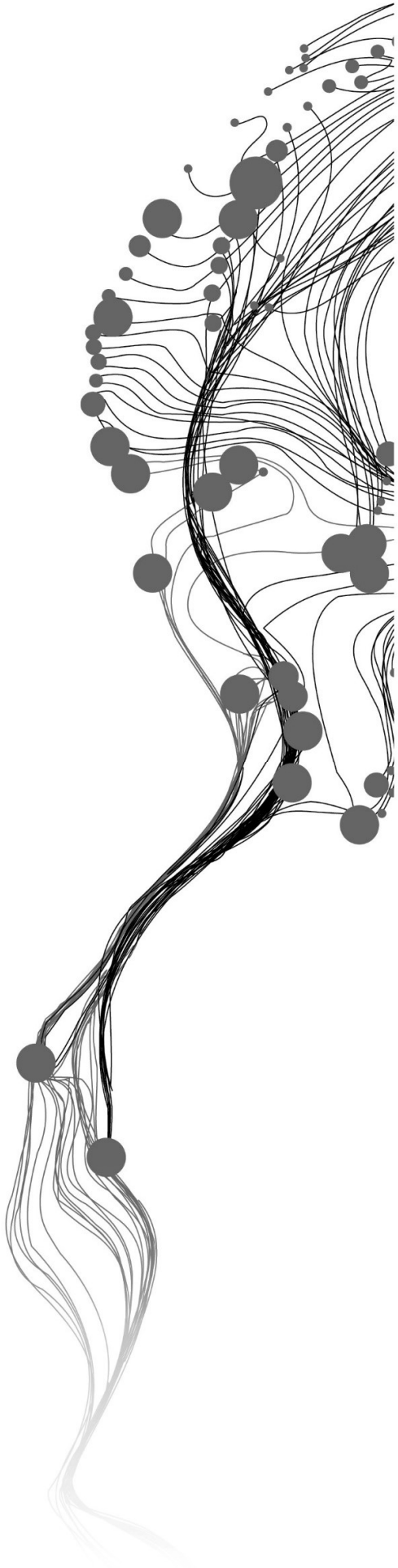
Exploring Pedestrian Navigation in Unfamiliar Urban Environments: Eye Fixation Analysis on Urbanscape Objects

AHMADREZA KARIM
August, 2023

SUPERVISORS:
dr. Gustavo García Chapeton,
dr. Franz-Benjamin Mocnik



Exploring Pedestrian Navigation in Unfamiliar Urban Environments: Eye Fixation Analysis on Urbanscape Objects

AHMADREZA KARIM
Enschede, The Netherlands, August, 2023

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Urban Planning and Management

SUPERVISORS:
Dr. Gustavo García Chapeton,
Dr. Franz-Benjamin Mocnik

THESIS ASSESSMENT BOARD:
Prof. Dr. Menno-Jan Kraak (Chair)
Dr. David Retchless (External Examiner, Texas A&M University at Galveston)

ABSTRACT

This study aims to analyze and understand pedestrians' fixation patterns on urbanscape objects while navigating unfamiliar urban environments, using eye-tracking technology. Thirteen participants engaged in a navigation task from the Basilica di San Lorenzo to Piazza della Signoria in Google Street View, and their eye movements were recorded and analyzed through Semantic Segment Anything (SSA). Three hypotheses were tested, focusing on correlations between dwell time, fixation duration, and deviation from the optimal route. The study revealed that buildings were the most observed and fixated objects across participants, serving as pivotal navigational guides. There was significant variation in fixation duration and count between participants who successfully completed the task and those who did not, indicating the importance of efficient scanning and rapid information processing. The methodology, which also included spatial consideration through Hausdorff distance and think-aloud data, offered a comprehensive understanding of visual behavior and navigation strategies. Findings contribute valuable insights into individualized navigation processes and provide practical recommendations for urban design and planning, emphasizing the importance of visually distinct building facades and pedestrian-friendly thoroughfares.

Keywords: Pedestrian Navigation, Eye-tracking, Urbanscape Objects, Urban Environments, Semantic Segmentation, Navigation Performance, Virtual Navigation

ACKNOWLEDGEMENTS

I would like to offer my heartfelt thanks to my supervisor, Paulo, quite possibly the coolest supervisor ever. His creative and innovative ideas guided me throughout this process, and his support was invaluable.

My gratitude also extends to my other supervisor, Gustavo. His constructive criticism continually pushed me to think of new possibilities, opening my mind to different perspectives.

A special thanks goes to my wonderful wife, Maryam, who was always there for me in tough situations, providing support no matter what. I couldn't have done this without her encouragement and love.

To my dearest friends, Morteza, Kimiya, and Sina, thank you for being such reliable companions. I knew I could always count on you for encouragement and insight.

Lastly, a high-five to myself for stepping out of my comfort zone and learning so much in such a short amount of time. This journey has been challenging, but I've grown so much because of it, and I'm grateful for the experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background and Context

The history of navigation goes back to the earliest recorded moments in humanity when our ancestors used fixed objects in the environment to find their way and acted as landmarks. For our hunter-gatherer ancestors, the core cognitive skill required for survival is the capacity to navigate (Yoder et al., 2011). The concept of a human sense of navigation has been discussed in the scientific literature for more than a century (Romanes & Darwin, 1884). We still use the navigation to meet our daily needs. Nowadays, urban trips are bound with navigation. Every day for doing a simple task in cities, we tend to plan for our travel and move toward the destination, which is called navigation.

Spatial disorientation, a term often used to describe a lack of awareness or confusion about one's location or direction, is not uncommon in urban environments (Benson, 2003). It can be a shared experience among city residents to feel momentarily confused and lost amidst shifting visual sceneries and vehicle motion that challenge the proper perception of direction (Gresty et al., 2008). However, this term has a more serious implication in certain contexts. For instance, 'spatial disorientation' is frequently used in more severe cases such as patients suffering from dementia or Alzheimer's disease, where the loss of spatial awareness is significant and persistent. For the average city dweller, on the other hand, this sense of disorientation is usually temporary and can be remedied by asking others for help or using navigation technologies.

Cities create a platform for daily activities; meanwhile, wayfinding for residents is inevitable. And lack of navigational ability can overshadow many aspects of everyday life (Aguirre & D'Esposito, 1999). Wayfinding in the city is a cognitive skill and is unique to every citizen in the city. Each person has an opinion about their navigating abilities, and humans use various navigational techniques (Shelton et al., 2013).

Eyesight is the primary and most crucial sense that aids in the visual perception of the environment. Daily navigation in humans comprises a combination of one or more different techniques, although visual information seems to predominate (Foo et al., 2005). The reflection of light in space and its entry into the eyeball leads to a set of biological processes in the eye and brain, which results in spatial understanding and interpretation of the relative position of our body and surrounding objects in the environment (Gibson, 1986). According to Ekstrom, (2015), the significance of the high-resolution visual representation of the human eye for how we navigate the environment should not be undervalued, even though it is somewhat helpful to conceive of our navigation system as including internal "maps."

Undoubtedly, navigation is a crucial part of the human experience. Whether we are thinking about how early people learned to survive in the wild or the more mundane concerns of how you buy groceries from the store, humans travel to their destination to meet their urban needs. According to Montello (2005), "*Navigation is coordinated and goal-directed movement through the environment by organisms or intelligent machines.*" In the same reference Montello (2005) describes, navigation as consisting of two main components: locomotion and wayfinding. The term "wayfinding" was established by Kevin Lynch in his book *Image of the City*, and after that, it was widely used and developed (Lynch, 1964). A comprehensive behavior for

seeking, exploring, and route planning from one area to another has been referred to as Wayfinding (Iftikhar et al., 2021).

Allen, (1999) proposes a functional distinction between wayfinding tasks. According to his paper depending on the goal, there are three main categories. Among these three types of tasks, quest is closely related to the subject of this research because he writes in his description: "*Quest involves travel from a familiar place of origin to an unfamiliar destination, a place which is known to exist but which the traveler has not visited previously.*" A quest is frequently led by route instructions, lists of landmarks, and tasks meant to lead from one to the next in order (Allen, 1999). One of the many definitions of a landmark is a building or object that designates a location and serves as a point of reference (Golledge, 1993). Landmarks can also assist in giving a moving agent a visual representation of an environment's essential elements, seen from a route viewpoint. This knowledge enables the moving agent to react appropriately in decision-making scenarios (Denis, 1997).

During the navigation process, landmarks play a crucial role in affecting the performance of wayfinding. The concept of a landmark encompasses any visual signal or object in the surroundings that can guide navigation. Within the scientific literature on spatial behavior, landmarks have been widely acknowledged as key elements that lead to the improvement of navigation (Caduff & Timpf, 2008). Whether natural formations or man-made structures, landmarks serve as reference points that help individuals orient themselves and navigate through various environments.

Virtual environments have become valuable tools for scientists studying navigation and behavior. These environments allow researchers to better understand and manipulate their surroundings, enabling them to address their inquiries more effectively. Conducting research in real-world settings presents various difficulties. Unlike studies conducted in controlled laboratory environments, it is challenging to exert influence and constraints on the real world. Ensuring identical conditions for all participants or controlling stimuli between participants to enhance task design is hard to achieve in real-world experiments (van der Ham et al., 2015). It is challenging to control potential disruptive factors in the real world, such as weather conditions, traffic, and noise (Burdea & Coiffet, 2003). The primary benefit of virtual environments is that they can be precisely modeled and controlled according to the specific requirements of an experiment, eliminating the need to construct a similar setup in the real world (Dombeck & Reiser, 2012).

While studies have been conducted in the fields of spatial and navigation capabilities, there remains a gap in quantitative analysis research within this area. The advancement of visual analysis technologies, such as eye-tracking, has substantially increased the capacity to analyze these phenomena. The primary contribution of this research is to identify the key urbanscape objects that attract attention during navigation in unfamiliar environments. Understanding these objects can lead to enhancements in urban environments to improve the navigation process for citizens. Moreover, this knowledge can assist urban designers in creating spaces with better visual comprehension or in making modifications to existing environments that may currently be overly complex and confusing for individuals.

1.2. Research Question and Objectives

The overall objective of this study is "to analyze and understand the fixation patterns of pedestrians on urbanscape objects while navigating themselves in an unfamiliar urban environment with a nearby mapped destination, without the aid of defined routes or navigational tools, and to assess the implications of these patterns on navigation performance." The study aims to provide insights and recommendations for urban design and planning based on these findings.

And the research questions are:

1. What are the urbanscape objects that pedestrians tend to fixate on during the process of self-navigation in an unfamiliar urban environment with a nearby mapped destination, and without defined routes or navigational tools?
2. What are the patterns in duration and frequency of fixations seen across pedestrians, how do they differ among individuals, and how do these patterns reflect the navigational behaviors of the participants?
3. What is the relationship between pedestrians' fixation patterns on urbanscape objects and their navigation performance in terms of successful reaching of the mapped destination?
4. Based on the understanding of the key urbanscape objects that attract pedestrians' visual attention during self-navigation, what insights and recommendations can be provided for urban design and planning?

1.3. Formulation of Hypotheses

In the following segment, the hypotheses that have been formulated for this study are presented. These hypotheses have been derived from the overarching research aim and serve to address the research question. In relation to this research question, the following three hypotheses are proposed:

Hypothesis 1: A significant correlation exists between the dwell time of Area of Interest (AOI) object among participants who successfully concluded the navigation task and those who failed to complete the task.

Hypothesis 2: There is a significant relationship between the duration of fixation on urbanscape objects by participants who successfully completed the navigation task and those who did not complete the task.

Hypothesis 3: There is a meaningful relationship between the fixation metrics of participants and the degree of deviation of their path from the optimal route.

These hypotheses propose specific relationships between the variables of interest in the study, and they are tested using the data collected from the participants. In the following sections, the methods used to collect and analyze this data will be described, and the results of the analysis will be presented to determine whether these hypotheses are supported by the data.

1.4. Thesis Organization

Chapter 1 of this thesis provides an overview of the background of the study and presents the aim and objectives to address the identified research problem. In Chapter 2, the focus is on the methods and technologies used to collect data about navigation and route selection. This includes an exploration of the tools and techniques that facilitate this process. Chapter 3, on the other hand, delves into the methodology, detailing the development of the proposed mixed-method approach and describing the specific process employed to gather data. In chapter 4, the results from the data collection are reported and then discussed in chapter 5. At last, chapter 6 presents the conclusions of this thesis, the ethical considerations, and the recommendations for future studies.

2. LITERATURE REVIEW

2.1. Introduction

The ability to navigate an unfamiliar urban environment is a crucial skill for individuals in today's increasingly mobile society. With the advent of virtual navigation tools such as Google Street View, individuals can now explore and familiarize themselves with new locations before even setting foot in them. However, little is known about how individuals use urban objects as visual cues while navigating in this virtual environment. This literature review aims to address this gap by providing an overview of current knowledge on the topic.

Specifically, this review will survey scholarly sources to identify relevant theories and methods related to the visual attention during virtual navigation. Additionally, this review will highlight gaps in the existing research and provide insights and recommendations for future studies. By synthesizing and critically evaluating the available literature, this review will provide a clear picture of the state of knowledge on the subject and lay the groundwork for further investigation.

Navigating an unfamiliar urban environment can be a challenging task. Pedestrians rely on various sources of information, such as maps, landmarks, signs, or GPS devices, to guide their navigation. However, the effectiveness and usability of these sources may vary depending on the characteristics of the environment and the individual preferences and abilities of the pedestrians (Fang et al., 2015).

Understanding how individuals use visual cues while navigating an unfamiliar urban environment is crucial for improving urban design and planning. The incorporation of well-placed visual cues, such as landmarks, signage, or distinct architectural features, can greatly enhance the intuitiveness and accessibility of a city's layout (Mohammadi Tahroodi & Ujang, 2021). This understanding allows urban planners and designers to create spaces that are more user-friendly and intuitive, catering to the natural instincts and preferences of pedestrians. By studying the relationship between visual cues and navigation, urban areas can be designed with pathways and landmarks that logically guide individuals, reduce confusion, and improve overall mobility and satisfaction (Cabanek et al., 2020). Furthermore, considerations for various demographics, including tourists or individuals with disabilities, can be taken into account, promoting inclusivity in design. This tailored approach, which relies on human-cantered design principles, not only makes cities more navigable but also contributes to creating vibrant, engaging, and liveable urban spaces.

2.2. Navigation and Wayfinding

Wayfinding, fundamentally a cognitive and problem-solving phenomenon, involves the complex cognitive process of utilizing various sources of information to navigate an unfamiliar environment. Understanding the theoretical foundations of wayfinding can lead to more meaningful and impactful design decisions for environments (Jamshidi & Pati, 2021). Studies such Liao et al., (2017) have explored the differences in visual attention in pedestrian navigation when using different types of maps or geo-browsers. Pedestrian navigation involves cognitive processes such as location awareness, orientation maintenance, destination recognition, and wayfinding (Farr et al., 2012). Hejtmánek et al., (2018) found that the amount of attention

individuals pay to GPS aids during navigation tasks has a significant impact on their spatial knowledge and navigation performance.

A study conducted by Bongiorno et al., (2021) suggests that, when navigating on foot, the human brain does not optimize for the calculation of the shortest possible route. Instead, individuals employ a vector-based navigation strategy, selecting paths that most directly point towards their destination, even if such paths are longer. This behavior has been observed in other species as well, ranging from insects to primates. The researchers propose that vector-based navigation, which requires less cognitive effort than the calculation of the shortest route, may have evolved to allow for the allocation of cognitive resources to other tasks.

Route choice

According to Golledge, (1999), "route choice" is a subprocess that can be performed on a mental representation of the spatial environment, or a "cognitive map," without requiring visual attention. When navigating in unfamiliar environments, individuals often rely on visual aids, such as maps or signs, to assist with route planning and choice. This was demonstrated in an eye-tracking study conducted by Netzel et al., (2017), which investigated how individuals use visual information to plan routes on metro maps. Similarly, Wiener et al., (2009) argued that wayfinding can be performed without prior knowledge of the environment through the use of visual cues and exploration. Meilinger et al., (2007) suggested that in map-based wayfinding, after planning and choosing a route, individuals transform and encode the map information into a mental representation for navigation. This involves remembering important spatial information to navigate successfully.

Orientation

According to Gunzelmann et al., (2004), orientation, which involves determining one's position relative to a reference frame, requires the integration of visual signals with knowledge about the environment. This process, also known as self-localization, as described by Kiefer et al., (2014), involves determining one's current position on a map, which is a crucial part of any wayfinding process. During self-localization, the wayfinder matches visually perceptible features of the environment, such as landmarks, with map symbols to constrain potential locations on the map. This process allows individuals to orient themselves in their environment and make informed decisions about their route. Eye-tracking studies have been used to investigate how individuals use visual attention during orientation tasks. Gunzelmann et al., (2004) found that the distribution of visual attention varies among individuals trained in different orientation strategies. This suggests that individuals may use different strategies to orient themselves in their environment, and that these strategies are reflected in their patterns of visual attention.
In an experiment on orientation, Peebles et al., (2007) used eye-tracking technology to investigate how individuals use visual attention during orientation tasks. They found that when presented with scenes containing salient 3D landmarks, participants' eye movements were strongly focused on those landmarks. This suggests that visual attention plays an important role in the orientation process, as individuals may use visual cues, such as landmarks, to orient themselves in their environment.

Virtual environments provide a valuable tool for studying navigation and wayfinding. Direction estimates in real environments, immersive virtual environments, and desktop virtual environments have been found to be more accurate and precise compared to other methods (Waller et al., 2004). Performance in real-life environments has been found to be superior to virtual environments for tasks that rely on survey

knowledge, such as pointing to the start and end points or drawing a map. However, performance in a hybrid environment, which combines real-world experience with simultaneous visual information on a tablet, did not significantly differ from real-life performance (van der Ham et al., 2015). Although, one crucial consideration is the complex nature of navigation, which poses challenges in analyzing various factors. Real-world environments are susceptible to uncontrollable factors such as weather conditions, traffic, and noise (Rey & Alcañiz, 2010). In this regard, virtual environments offer greater control over the environmental conditions, presenting more opportunities for manipulation and analysis. Utilizing geospatial Google Street View in a VR setting has been found to enhance students' motivation for spatial knowledge acquisition and provide a valuable educational tool for spatial training (Carbonell-Carrera & Saorín, 2017).

The use of virtual environments can provide insights into real-world navigation behavior. There are powerful analogies between movement in virtual environments and movement in real environments (Dalton, 2001). Studying movement in immersive virtual environments can provide insights into the micro-scale decision-making processes that contribute to emergent regularities observed in real-world pedestrian movement. However, some methods used to gather information about navigation in virtual environments have their limitations. For example, the quality of Think-Aloud protocols depends on the abilities of the participants and may only provide limited information for landmark identification (Viaene et al., 2014).

2.3. Landmarks and Their Impact on Navigation Performance

Landmarks play a crucial role in pedestrian navigation (Chan et al., 2012). examine the widespread application of the landmark concept, which now encompasses any visual stimulus in an environment with the potential to affect navigation. By proposing that landmarks extend beyond mere objects, Chan et al. consider the interactions of landmarks with their surroundings and other elements in the environment. This means examining how landmarks relate to and influence other features, such as roads, paths, or buildings, and how they collectively guide or influence a person's navigation choices. These interactions can include how landmarks draw attention, how they are positioned relative to other cues, and how they may align with cultural or societal meanings in a given context.

Yesiltepe et al., (2021) present a comprehensive review of the literature regarding the selection of landmarks in wayfinding, focusing primarily on large-scale urban environments and outdoor settings. The review centers on two crucial aspects of landmarks: their **visibility** and **salience**. In navigation, literature highlights visibility and salience as key characteristics of landmarks. Visibility is defined as the capability of a landmark to capture the observer's eye or attention, whereas salience is understood as the standing out or uniqueness of a landmark in its environment (Li et al., 2017). The concept of landmark salience is complex, stemming from the observer's physical and mental perspective, the surrounding environment, and the objects within it. Salience is further described through a three-part Saliency Vector, consisting of Perceptual Salience, Cognitive Salience, and Contextual Salience. This distinction, considering both voluntary and involuntary focus of visual attention in relation to the context, lays the groundwork for a framework that explains the interplay between the observer, the environment, and the landmark (Caduff & Timpf, 2008). According to the findings, there is general agreement concerning the significance of landmark location. Landmarks situated along the route and at decision points (where a turn is required) prove to be more effective in aiding wayfinding tasks.

The use of landmarks has been shown to decrease navigational errors and optimize navigation performance for pedestrians (Rehrl et al., 2010). Participants have been observed to use recognizable and contextually relevant object landmarks, such as 3D models of everyday objects, when they are present. These landmarks often enhance navigation due to their clear connection to the surrounding environment or cultural significance. In contrast, participants have difficulty using less informative landmarks, such as colored abstract paintings, as aids for successful navigation (Hamid et al., 2010). These types of landmarks might lack clear relevance to the setting or fail to resonate with the observer's understanding or experience of the space. This distinction illustrates the importance of choosing appropriate landmarks that align with the needs and comprehension of those navigating through the environment.

The effectiveness of landmarks in navigation can be influenced by the way they are presented. When comparing 2D and 3D electronic maps, there are significant differences in fixation time and saccade amplitude. Users tend to have shorter fixations and larger saccades in 2D maps, while longer fixations and smaller saccades occur in 3D maps (Lei et al., 2016). Incorporating landmarks in 3D representations can improve the usability of pedestrian navigation systems, particularly in aiding decision-making at complex locations (Lei et al., 2016). Landmarks that are focused on longer and more frequently during navigation transfer onto the mental map, suggesting that paying more attention to specific landmarks enhances their imprint on the cognitive map (Franke & Schweikart, 2017).

2.4. The Role of Eye-Tracking in Navigational Research

Eye-tracking technology provides valuable insights into visual attention during navigation. An eye tracker is a device for measuring eye positions and eye movement. Infrared eye trackers work by emitting a near-infrared (NIR) light beam towards the center of the eyes (pupil). This light is reflected in the user's eyes, causing detectable reflections in both the pupil and the cornea. The reflections are captured by the eye tracker's cameras, and through filtering and triangulation, the eye tracker determines where the user is looking and calculates eye movements data (Ware & Mikaelian, 1986).

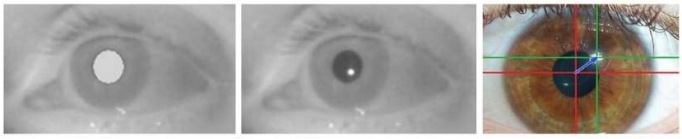


Figure 1 Corneal reflections (CR) are generated by utilizing the center of the pupil and infrared/near-infrared non-collimated light in eye-trackers (Holmqvist et al., 2012).

New eye trackers can measure various variables and types of eye movements. The oculomotor system controls how our eyes move. It uses parts of the visual and vestibular systems. It manages different eye movements like quick jumps, smooth tracking, bringing eyes together, and reflexes for balance (Robinson, 1968). In their study Mahanama et al., (2022) provide a comprehensive overview of the primary oculomotor events and their quantifiable characteristics. Moreover, they help us understand different eye

movements by introducing various widely used techniques for analyzing eye tracking, as illustrated in Figure 2.

Eye movement data can be analyzed in terms of fixations and saccades, where fixations represent periods of visual gaze fixated on a specific location, and saccades are rapid eye movements occurring between consecutive fixations (Fischer & Weber, 1993). The definitions of fixation, saccade, and smooth pursuit are detailed in Table 1.

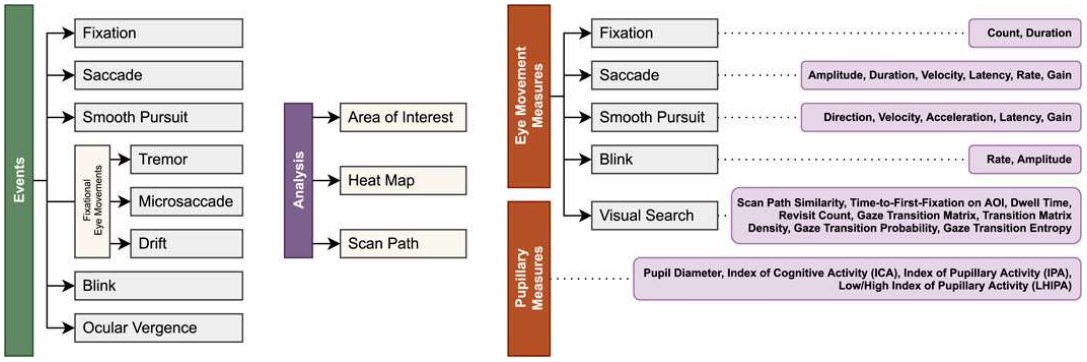


Figure 2 Summary of oculomotor events provided by (Mahanama et al., 2022)

Fixations	Saccades	Smooth pursuit
Fixations are periods during which the eye remains relatively still, enabling the brain to process visual information from the fovea, the area of the retina responsible for sharp central vision. These periods can last from 200 milliseconds to 2 or 3 seconds.	Saccades are rapid eye movements that occur when the eye jumps from one fixation point to another. Saccades are the fastest movements produced by the human body and can last anywhere from 20 to 200 milliseconds.	Smooth Pursuit movements allow the eyes to smoothly follow a moving object. These movements are slower than saccades and are used to track objects in motion.

Table 1 Definition of different eye movements (Hutton, 2020)

Eye-tracking technology has been used to understand the effectiveness of commonly used navigational elements in interface design (Ford et al., 2020). By analyzing eye tracking data, researchers can identify areas of interest, patterns of visual attention, and potential usability issues related to navigation. This information can inform the optimization of navigational elements to improve user experience. Furthermore, eye-tracking emulation software can be utilized by urban designers and architects to understand how humans unconsciously respond to visual stimuli in the built environment (Hollander et al., 2021). This tool can help them assess the effectiveness of design elements and make informed decisions during the design process.

Virtual environments provide a valuable tool for studying visual attention during navigation. Our attentional system can prioritize and process information related to objects even when they are partially occluded or defined by subjective contours (Moore et al., 1998). The relationship between gaze behavior and target objects has been found to be statistically significant across different target conditions (Enders et al., 2021). There is also a significant relationship between gaze behavior and the identification of target

objects during a virtual navigation task (Enders et al., 2021). These findings emphasize the utility of virtual environments for studying active visual search and provide insights into the dynamics of gaze behavior during navigation and visual search tasks in realistic virtual environments.

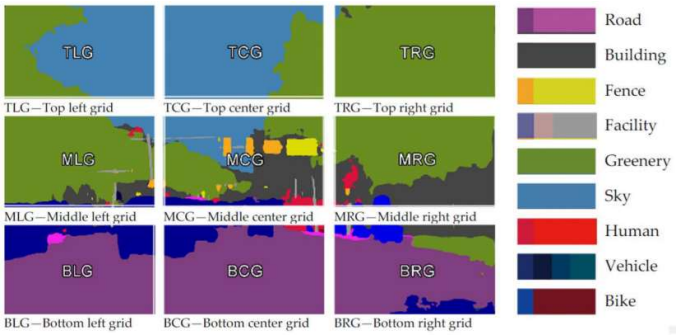


Figure 3 Using semantic segmentation, the study conducted by (Yue et al., 2022) revealed that participants consistently directed their attention towards the frame center, irrespective of the mode of transportation.

Individual differences have been observed in visual attention during navigation. Sex differences in navigational behavior have been observed only in environments without landmarks. In environments with multiple landmarks, the sex differences disappeared (Andersen et al., 2012). Females exhibited sustained gaze towards landmarks throughout the task, while men's gaze towards landmarks decreased over time. Males and females have also been found to show differences in the use of distal cues during a virtual environment navigation task. Females rely more on landmark information, while males are more likely to utilize both landmark and geometric information (Sandstrom et al., 1998). Age differences have also been observed in spatial memory and navigational behavior in a virtual environment task. Older participants performed worse compared to younger participants in terms of solving each trial, distance traveled, and spatial memory errors (Moffat et al., 1998).

2.5. Summary

In conclusion, this chapter provides a background of the thesis. In the literature suggests that understanding how individuals use visual cues while navigating an unfamiliar urban environment is crucial for improving urban design and planning. Eye-tracking technology has been used to understand the effectiveness of commonly used navigational elements in interface design and to inform the optimization of navigational elements to improve user experience. The use of landmarks has been shown to decrease navigational errors and optimize navigation performance for pedestrians. Virtual environments have been found to provide valuable insights into the micro-scale decision-making processes that contribute to emergent regularities observed in real-world pedestrian movement. Utilizing geospatial Google Street View in a VR setting has been found to enhance students' motivation for spatial knowledge acquisition and provide a valuable educational tool for spatial training. Individual differences, such as sex and age, have been observed in navigational behavior and gaze patterns. These findings have important implications for urban design and planning, as they highlight the importance of incorporating landmarks in the design of pedestrian navigation systems. Considering landmarks in route choice models for pedestrian movement simulation improves the realism of the model and enables pedestrians to make use of relevant urban information during navigation.

3. METHODOLOGY

3.1. Introduction

This section details the methodology employed in the study. A multi-faceted approach was designed, incorporating both quantitative and qualitative data analysis techniques, to investigate complex interactions within the chosen field of study. The ensuing subsections elaborate on the specific methods used, including the participants, tools, procedures, and the rationale behind these choices.

In order to simulate the experience of navigating an unfamiliar urban environment using Google Street View, a group of 13 participants was recruited to perform a navigational task in the city center of Florence. Various types of data were collected from each participant during this task. Eye-tracking data was recorded to analyze participants' visual attention patterns as they navigated the virtual environment. Screen captures of participants' performance in Google Street View were also captured to provide a detailed account of their navigation behavior. Furthermore, data regarding the paths taken by participants to complete the task and their verbalized thoughts during the task were collected as think-aloud protocol.

To facilitate a more concise quantitative analysis, the video recordings of participants' performance in Google Street View were subjected to semantic segmentation using the Semantic Segmentation Anything model. This allowed for an examination of the coordination between gaze patterns, as indicated by eye-tracking data, and the objects present in each frame of the video.

The collected data was analyzed through several approaches. Initially, eye-tracking data was utilized to identify the specific urban objects that participants fixated on during their navigation. The duration and frequency of these fixations were then analyzed to comprehend how visual attention was distributed among different objects. Subsequently, an investigation was conducted into the relationship between participants' fixation patterns and their navigation performance, encompassing measures of successful completion of the task.

By integrating these diverse data types and analyses, this study aims to provide an understanding of how individuals employ urban objects as visual cues when navigating an unfamiliar urban environment. Google Street View is used as a simulation tool to recreate real-world conditions, allowing for a more controlled examination of navigation behavior. The insights gained from this research will be valuable for urban design and planning, as well as the development of virtual navigation tools, by offering a deeper understanding of how people navigate in real life.

The overall steps of the methodology can be seen in the Figure 4.

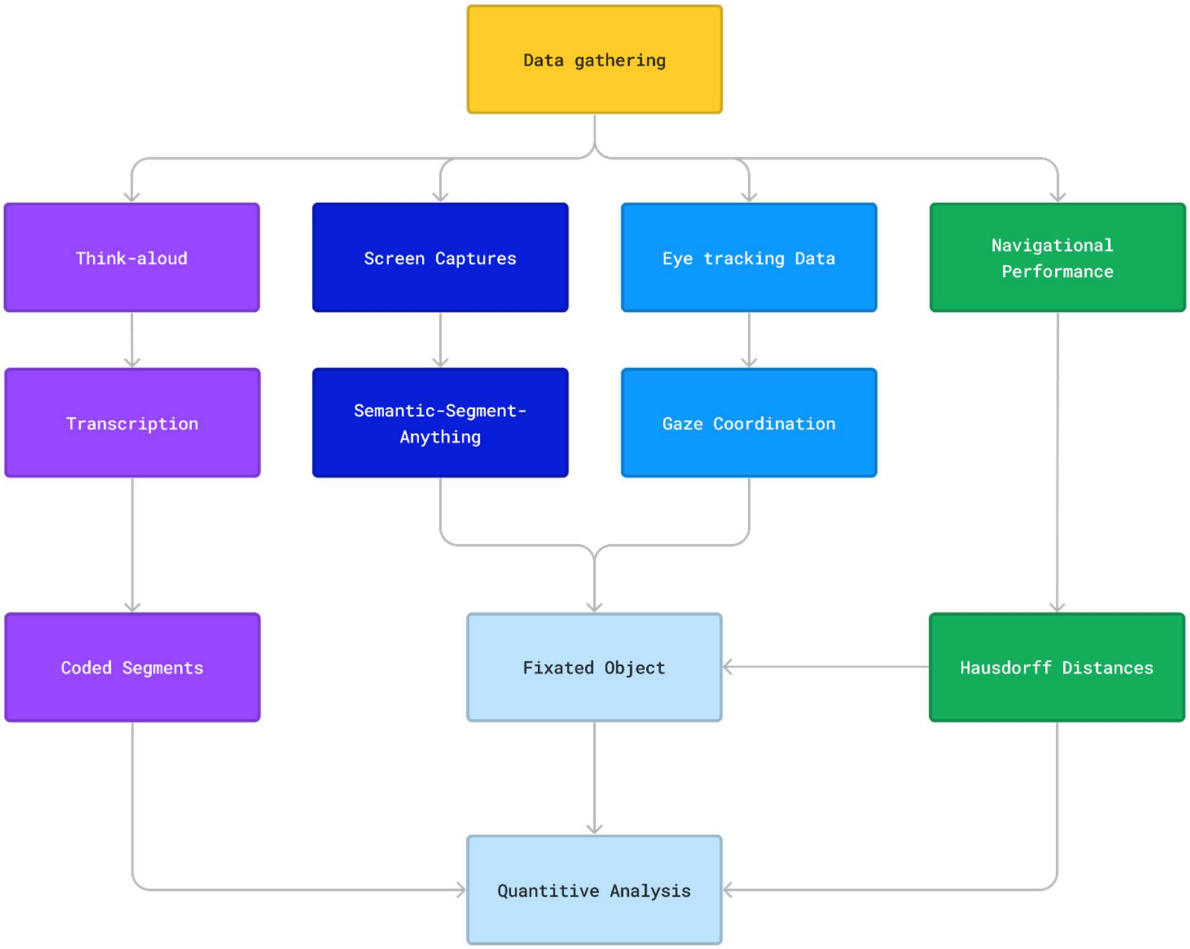


Figure 4 Overall Methodology

3.1.1. Selection of the Study area

The selection of the study area required careful consideration due to its significant impact on individuals' navigation in different environments and surroundings. Several criteria were considered when choosing the study area, recognizing that different environments might present unique challenges or features, even though the essential navigational task remains the same: going from point A to point B. While the fundamental objective of navigation does not change, the study area's specific characteristics can influence factors such as complexity, available cues, and potential obstacles, all of which may affect the outcomes of the research.

Start to End distance

The choice of starting and ending distances within the study area was determined by considering two essential limitations that influenced the selection of urban scenes. Firstly, the distance between the start and end points played a key role in determining the complexity of the navigational task, as a longer distance generally increases the level of difficulty. Secondly, there was a time constraint to consider, as participants' travel time would significantly increase if the start and end points were too far apart. To address these limitations, a meticulous selection process was undertaken to filter urban scenes based on an appropriate distance range, striking a balance between navigational complexity and time efficiency.

Moving direction

To introduce an element of unpredictability and enhance the complexity of the navigation task, it is recommended to select a route that follows a diagonal path within a rectangular area, from one corner to the diagonally opposite corner. Choosing two adjacent corners would create a simpler and more predictable route. By incorporating a diagonal moving direction, the navigation task becomes more intricate, leading to more diverse outcomes (Brunyé et al., 2015).

Minimal Topological Variation

The absence of topographical variation in the study area has implications for the navigational task. When the city exhibits a flat topography with minimal changes in altitude, participants are not faced with significant variations in building heights or terrain. This flat terrain ensures that landmarks and objects within the city are generally at the same level, making their selection and visibility more consistent. In contrast, if the city had varied topography, with some buildings or objects at higher elevations, it could potentially impact the selection and visibility of landmarks. Therefore, by intentionally choosing a study area with minimal topographical variation, the influence of altitude-related factors on landmark selection is reduced, allowing for a more controlled and standardized navigational experience (Brunyé et al., 2015).

Landmark guidance

To simplify the navigational task for individuals with varying spatial abilities, it is crucial to select a study area that includes landmarks. These landmarks serve as essential cues to support participants throughout the task (Rehrl et al., 2010). Additionally, having a visual connection between the first and second landmarks is advantageous, as it allows for easy spotting of the second landmark from the vicinity of the first landmark without significant effort. This visual connection facilitates seamless navigation and aids participants in maintaining their sense of direction. Moreover, the inclusion of a middle landmark serves as a reorienting point within the study area, further assisting participants in navigation.

City morphology diversity

City morphology diversity is crucial to ensure that the navigation process is not entirely predictable and devoid of challenges. When the city's layout offers a diverse range of features, such as varying street patterns or irregular shapes, it introduces an element of uncertainty and requires participants to adapt their navigation strategies. In contrast, cities with a checkered grid or completely straight streets tend to make the navigation process highly predictable, thereby diminishing the overall challenge.

The selection of the study area was conducted based on the criteria discussed above, aiming to create a suitable environment for investigating navigational processes. An urban area in the city of Florence, Italy, was chosen as the study site. While many cities around the world could have fulfilled the criteria, Florence, renowned for its rich history and architectural landmarks, was selected to offer a specific setting to explore human navigation.

Figure 5 The Cathedral of Florence, prominent landmark visible from various points across the city. (Google LLC, 022)

Figure 6 provides a visual representation of the selected geographical area and the designated starting and ending points for the navigational task. The starting point was situated in front of Basilica di San Lorenzo (43.7755°N, 11.2536°E), and the endpoint was located in front of Piazza della Signoria (43.7696°N, 11.2558°E). The study area encompasses a diverse range of urban features, including bustling streets, iconic landmarks, and various cityscapes. The starting and ending points for the navigational task were deliberately determined to create a challenging yet achievable task. By choosing a diagonal path within the study area, the task's complexity was enhanced, as this route required participants to navigate through an intricate urban layout without the assistance of maps.

Florence's unique combination of architectural landmarks and diverse street patterns added to the navigational challenge, requiring participants to rely on visual cues and their spatial abilities. This complexity reflects the real-world challenges that individuals may encounter when navigating unfamiliar urban environments, making the study relevant and robust. The selection of these points takes into account the desired distance range, allowing participants to navigate a substantial yet manageable distance of approximately 750 meters along the optimal route. The chosen points also consider the presence of significant landmarks and environmental cues that can aid participants in their navigation. By carefully planning the starting and ending points, the study aims to create a challenging yet rewarding navigational experience.

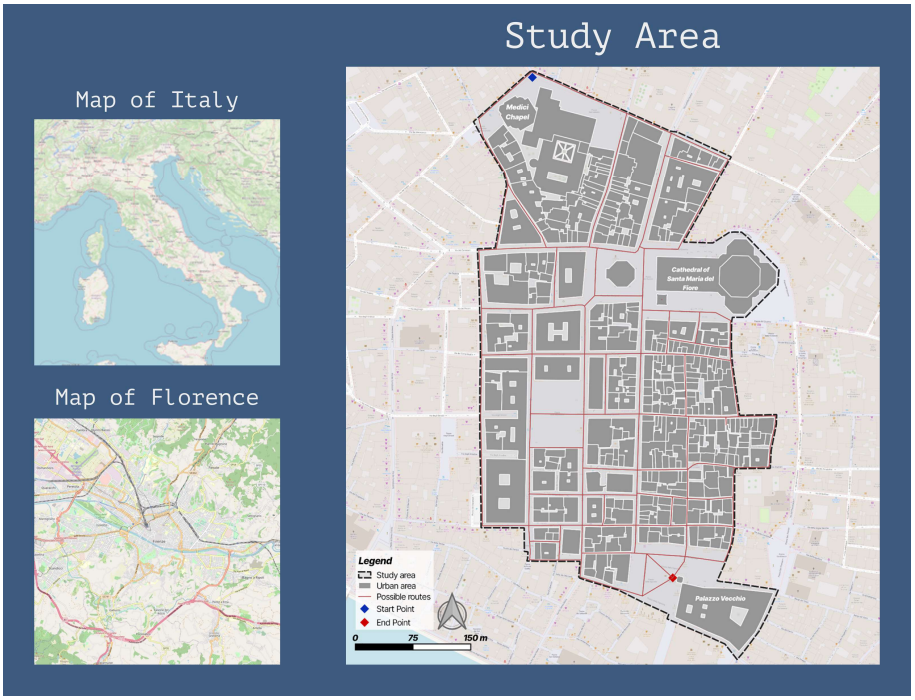


Figure 6 The study area from the top left Italy scale to Florence and finally city center of Florence.

3.2. Participants

3.2.1. Recruitment Process and Sample Size

Participants for the navigational survey were recruited from the student population at the University of Twente, using a direct approach. Students were approached in person and invited to participate in the study. The invitation included information about the purpose of the study, the time commitment required, and a gesture of gratitude in the form of a chocolate offered for participation.

The sample size of 13 participants for the navigational task was determined based on a combination of factors, including the complexity of the task, the time required for this task, and the need for a manageable data set for analysis. By comparing with similar studies in the field and conducting a statistical power analysis, it was concluded that a sample of 13 would allow for meaningful insights while maintaining the quality of the research process. This size also aligns with the resources available for the study, providing a balance that supports the overall objectives of the research.

Sample sizes similar to the one employed in this research have been successfully utilized in studies within the field of cognitive psychology and navigation. For instance, research investigating eye tracking often involves participant groups ranging from 13 to 50 individuals (Winkler & Subramanian, 2013), while investigations exploring the influence of landmarks on wayfinding behavior have used groups ranging from 10 to 15 individuals (Ruddle et al., 1997). These examples demonstrate that the selected sample size is within a range that has proven capable of detecting significant effects and generating meaningful insights into human navigation processes.

3.2.2. Inclusion and Exclusion Criteria

Prospective participants were required to meet two key criteria for the study:

1. They must be at least 18 years old, the legal age for participation in research studies.
2. They must have had no prior experience with the study area.
3. They must have had no visual impairments that could potentially impact the results of the eye-tracking assessment.
4. They must possess the ability to communicate in English, ensuring a common language for instructions and data collection.

Individuals with visual impairments were excluded to prevent any potential biases or distortions in the eye-tracking results. This careful selection process aimed to maintain the integrity and accuracy of the data collected during the navigational task, effectively minimizing potential biases and confounding factors. By ensuring a consistent group of participants with similar levels of unfamiliarity and accurately measurable eye movements, the study aimed to preserve the reliability and validity of its findings.

3.2.3. Participant Demographic Information

The demographic characteristics of the study participants are summarized as follows. A total of 13 participants took part in the study, consisting of 8 males and 5 females. The age of the participants ranged from 24 to 35 years (Figure 7), with an average age of 29.30 years and a standard deviation of 3.038. The median age, which represents the middle value in the dataset, was found to be 30 years.

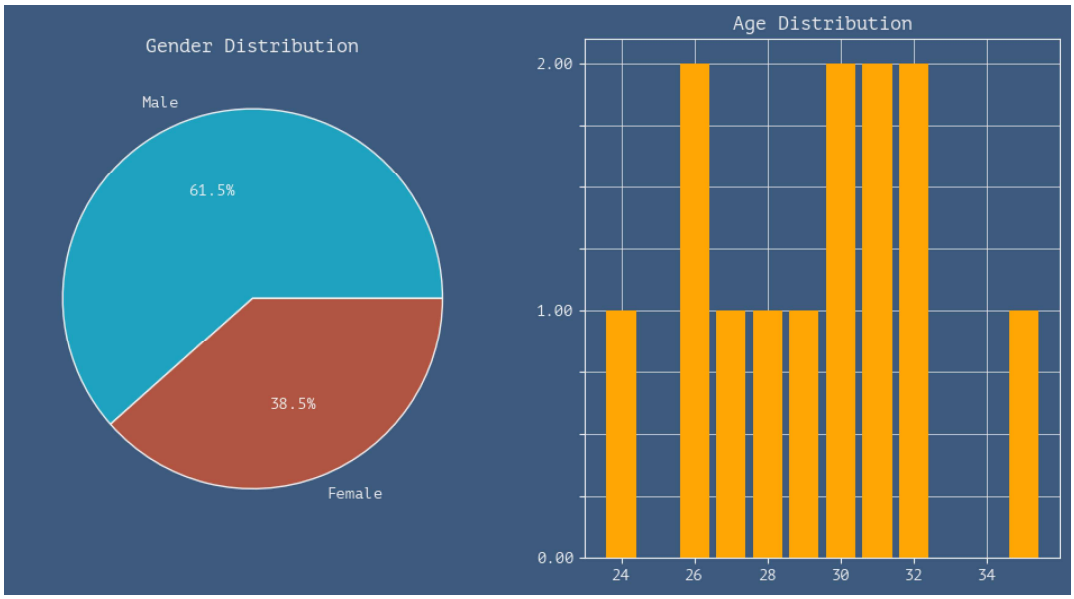


Figure 7 Demographic Distribution of Participants

The demographic data of the participants reveals a mean age of approximately 29.3 years, indicating that the majority of the participants fall within the young adult to adult age range. The standard deviation of approximately 3.04 years suggests a relatively close grouping of ages around this mean value, signifying

less variability in the participants' ages. The minimum and maximum ages of the participants are 24 and 35 years, respectively, denoting that all participants fall within this age bracket.

In terms of gender distribution, the dataset comprises more male (8 in total) than female participants (5 in total). Thus, the sample is not evenly split between the two genders, with males making up a larger proportion of the participants.

3.3. Data Collection

3.3.1. The Virtual navigation task

The navigational task was conducted at the VISUSE Lab, in the Geo-Information Processing Department of the University of Twente, where a well-lit and comfortable environment was prepared for the participants. Prior to the survey, participants were greeted, and the researcher introduced himself. A comprehensive overview of the study's objective, which focused on exploring individuals' navigation behavior in unfamiliar urban environments, was provided. The implementation of eye-tracking technology was explained, emphasizing its crucial role in monitoring participants' eye movements during the task. The task was deemed successfully completed when participants identified the endpoint, Piazza della Signoria, and positioned themselves in close proximity to the piazza.

Figure 8 Screenshots captured from the video. 1 - Shows the location of the first landmark on the map; 2, 3, 4 - Show examples of the 3D projection of each landmark in the video.

To facilitate participants' understanding of the task, an informative introductory video was created. This video zoomed in from a large-scale view to Florence and eventually study area. After that it represents the start and end points on a 2D map. The video also highlighted the location of each landmark and their sequential connection, represented by a dash lines, to guide participants towards the endpoint.

Furthermore, 3D rotating views of each landmark were presented, enabling participants to observe the landmarks from different angles and develop a cognitive map, as investigated by Lei et al., (2016). Figure 8 displays a segment of the video related to the presentation of the landmarks. In this part of the video, the researcher provided three key physical characteristics for each landmark to enhance participants' recognition.

Basilica di San Lorenzo: Medium dome, brick walls, and a long wall attached to the building.
Florence Cathedral: White marble building, a prominent dome visible from various locations, and a large adjacent tower.
Piazza della Signoria: A tower attached to a building, brick exterior, and a plaza in front of it, serving as the final point.
Upon introducing each landmark, the video concluded. (The video can be accessed via this link: https://youtu.be/6N2Q61uJo3g)

The task was executed using the Google Street View API, hosted on a local server facilitated by Python. Utilizing the API rather than the Google Street View website provided several advantages that contributed to the study's success. First, the API removed distracting elements, such as shop's labels and other features that exist in the Google Street Map but were irrelevant to the navigational task at hand. This decluttered the environment, allowing participants to focus solely on relevant objects and environments, as shown in Figure 9.

Moreover, the API offered a streamlined navigation experience with fewer images, making the task more straightforward for participants. The reduced image load minimized potential loading delays and movement-related bugs, resulting in smoother and more seamless navigation.

In addition to these benefits, the API supported participants' navigation experience by incorporating arrows on the ground, indicating permissible directions. This feature was helpful, assisting participants in making more informed decisions about their routes and destinations.

Lastly, the use of the API ensured a standardized starting view for all participants, a task that could also be accomplished through the website but was more easily programmable using the API. This consistency in starting views contributed to enhancing the study's reliability and comparability of results.

Figure 9 Difference between original Google Street View and its API in terms of removing minimap and shop's labels

Participants were provided with clear instructions for navigation, with the option to use a mouse or keyboard for movement within the virtual environment. In cases where images in Google Street View did not function optimally, alternative methods such as arrow keys or clicking on arrow-shaped markers displayed on the pavement were recommended. Participants were also advised to be mindful of image rotations between each frame and adjust their movements accordingly to maintain accurate spatial orientation. To ensure data integrity, participants were told not to use the mouse scroll function, as it could introduce errors.

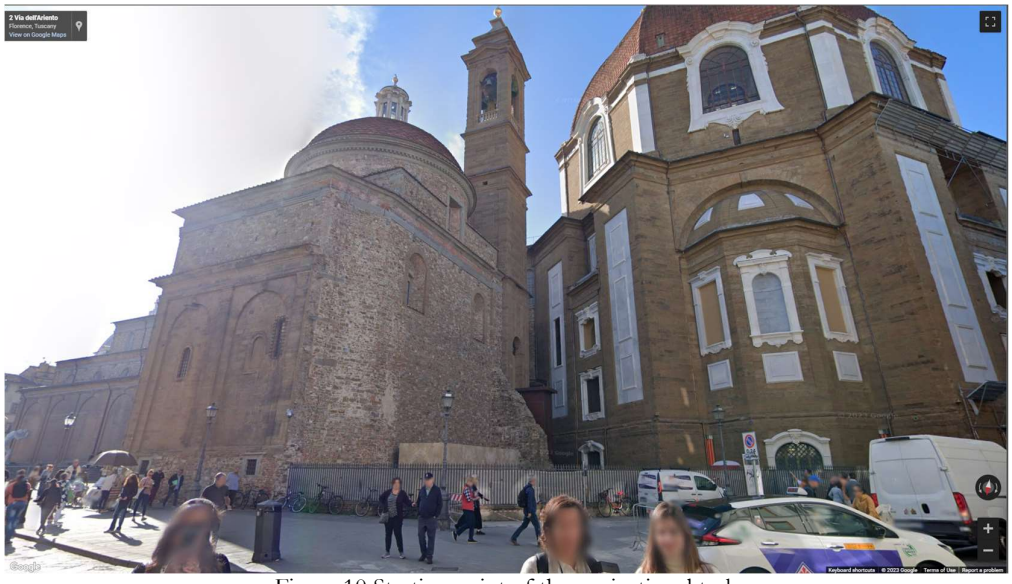


Figure 10 Starting point of the navigational task

Participants were explicitly informed that the navigational task had no predetermined time limit to avoid any influence on their behavior or a rushed approach. The duration of the task was left open-ended, allowing participants to complete it at their own pace.

However, certain conditions were established under which the researcher retained the right to conclude the task. If a participant was significantly distant from reaching the designated endpoint, if the task exceeded 20 minutes (although this time limit was not disclosed explicitly), or if the participant expressed a sense of being completely lost and devoid of hope in finding their way, the researcher had the option to terminate the task. These conditions ensured the overall feasibility and effectiveness of the study while prioritizing participant comfort and engagement.

As an expression of gratitude for their participation, each individual participant received a small token of appreciation in the form of a chocolate. This gesture served as a symbolic conclusion to the survey, and participants were also clearly informed when their involvement in the study was officially over.

3.3.2. Eye-tracking Data Collection

Prior to the task commencement, participants were provided with detailed instructions regarding the setup and calibration procedures for the eye-tracking system. The purpose of these instructions was to ensure optimal data collection and accurate interpretation of participants' eye movements. The following sections offer further insights into the calibration process, the significance of calibration, and the specific instructions given to participants.

The calibration process holds vital importance in eye-tracking research as it establishes a precise mapping between the participant's gaze and the screen coordinates. By calibrating the eye-tracking system, the accurate determination of participants' gaze positions on the screen throughout the task was made possible. This facilitated the analysis of visual attention and gaze patterns with enhanced accuracy.

Participants were guided to sit comfortably in front of the eye tracker, with an emphasis on maintaining a steady posture and minimizing unnecessary movement. The provision of a stable chair aimed to ensure a consistent viewing angle and reduce potential sources of error that could impact eye tracking data. Additionally, special attention was given to the setup to ensure that participants maintained a comfortable posture at an approximate distance of 60 cm from the screen. This specific distance was chosen for its significance in calculating the eye-tracking fixation metric based on the I-DT (Identification by Dispersion-Threshold) algorithm, which will be discussed further in the next sections. Maintaining this distance is essential to optimize the accuracy of the eye-tracking measurements and enable precise fixation analysis using the designated algorithm, contributing to the consistency and validity of the study's findings

The assessment of each participant's posture and the eye tracker's detection of their eyes was carefully conducted using the "Tobii Pro Eye Tracker Manager" software. This evaluation ensured optimal positioning and alignment between the participant's eyes and the eye-tracking system. When necessary, slight adjustments were made to the participant's distance from the sensor to maximize the accuracy of eye tracking measurements.

Detailed step-by-step instructions on the calibration process were provided to participants. They were instructed to focus their gaze on a moving dot displayed on the screen. Participants were asked to follow the movement of the dot or fixate on each point as it appeared. This calibration procedure enabled the eye-tracking system to establish an accurate and personalized gaze mapping for each participant.

Throughout the calibration process, the researcher closely monitored the calibration quality to ensure precise gaze mapping. If the quality was found to be inadequate or any discrepancies were detected, recalibration was performed to enhance the precision of eye-tracking data. This attention to calibration quality was essential to identify any potential tracking errors or artifacts, thereby ensuring the reliability and accuracy of the collected data.

These comprehensive measures and instructions were implemented to optimize the eye-tracking setup and ensure accurate and reliable collection of eye tracking data. By carefully calibrating the eye-tracking system and providing participants with clear instructions, the aim was to minimize sources of error and enhance the validity of the findings.

3.3.2.1. Apparatus

The apparatus for this navigational task included several key components to facilitate accurate tracking and a seamless experience for the participants. The core of the setup was the Tobii Pro Fusion eye tracker, which provided insights into the participants' gaze patterns with a sample rate of 120 Hz. Participants performed the task on a Dell Precision 3561 laptop, equipped with a 15.6-inch display monitor that had a screen resolution of 1920 × 1080 pixels. This resolution was chosen to present the map stimuli with high visual fidelity. To record participants' think-aloud data, a microphone was also used as part of the apparatus. These tools collectively contributed to the successful execution of the navigational task, allowing for precise data collection and a comprehensive understanding of the participants' navigation strategies.

3.3.2.2. Video Data Extraction

Upon completion of the navigation task by each participant, video screen recordings of their performances were extracted from the Tobii Pro Lab software. The beginning portion of each video, covering calibration, was removed as it wasn't connected to the navigation task. The recording process was stopped when the third landmark was recognized by the participant.

The duration required for each participant to complete the task varied, as individual spatial abilities to navigate their path differed. The following table depicts the duration of the video primarily reflecting the time each participant took to conclude the task:

The Tobii Pro Lab software's Event feature was employed to indicate these two points on the video. This function not only made the removal of the calibration phase possible but also marked two critical moments, known as "Start_survey" and "End_survey", on the eye-tracking data.

Further analysis was conducted with these refined videos. The videos were transformed into MP3 files and the spoken words were transcribed, creating a set of think-aloud data. Each video was split into individual frames, which were then fed through a semantic segmentation model. The sequence of selected images in the video allowed for an understanding of the route each participant took.

3.3.3. Think-Aloud Protocol

Think-aloud Protocols and Cognitive Interviews are structured methodologies employed to document and analyze concurrent cognitive processes. In qualitative research using the think-aloud technique, participants are requested to verbalize any thoughts, feelings, and actions, providing an insight into their metacognitive processes. This approach allows investigators to interpret cognitive operations and performance, thereby aiding in the exploration of decision-making and other mental processes (Wolcott & Lobczowski, 2021).

Think-aloud protocols entail the articulation of thought processes during cognitive tasks such as reading, problem-solving, among others (Oster, 2001). Participants may voice comments, pose questions, formulate hypotheses, or infer conclusions (Trapsilo, 2016).

In this research, the think-aloud protocol was employed to gather data from participants, prompting them to verbally express their thoughts as they navigated through the task. A clear explanation and examples of how to think aloud were provided prior to the commencement of the survey. During the navigation task, prompts were used to remind participants to continue verbalizing their thoughts. This approach ensured that insights into the participants' thought processes were captured without significantly disrupting their concentration. Examples of these prompts can be found in Annex 1 of the participation survey protocol.

Upon collecting data from all participants, the audio from the video recordings was extracted. This collection of audio data, which originated from the participants' think-aloud comments during the task, can be further analyzed.

3.4. Data Processing

The following section offers a introduction to the data processing steps used in this study. We started with raw data and performed multiple procedures to clean, verify, and extract meaningful information. This involved both manual checks and automated techniques. Detailed information about these processes is provided in the later sections. Also a detailed version of steps taken can be found in the Figure 11.

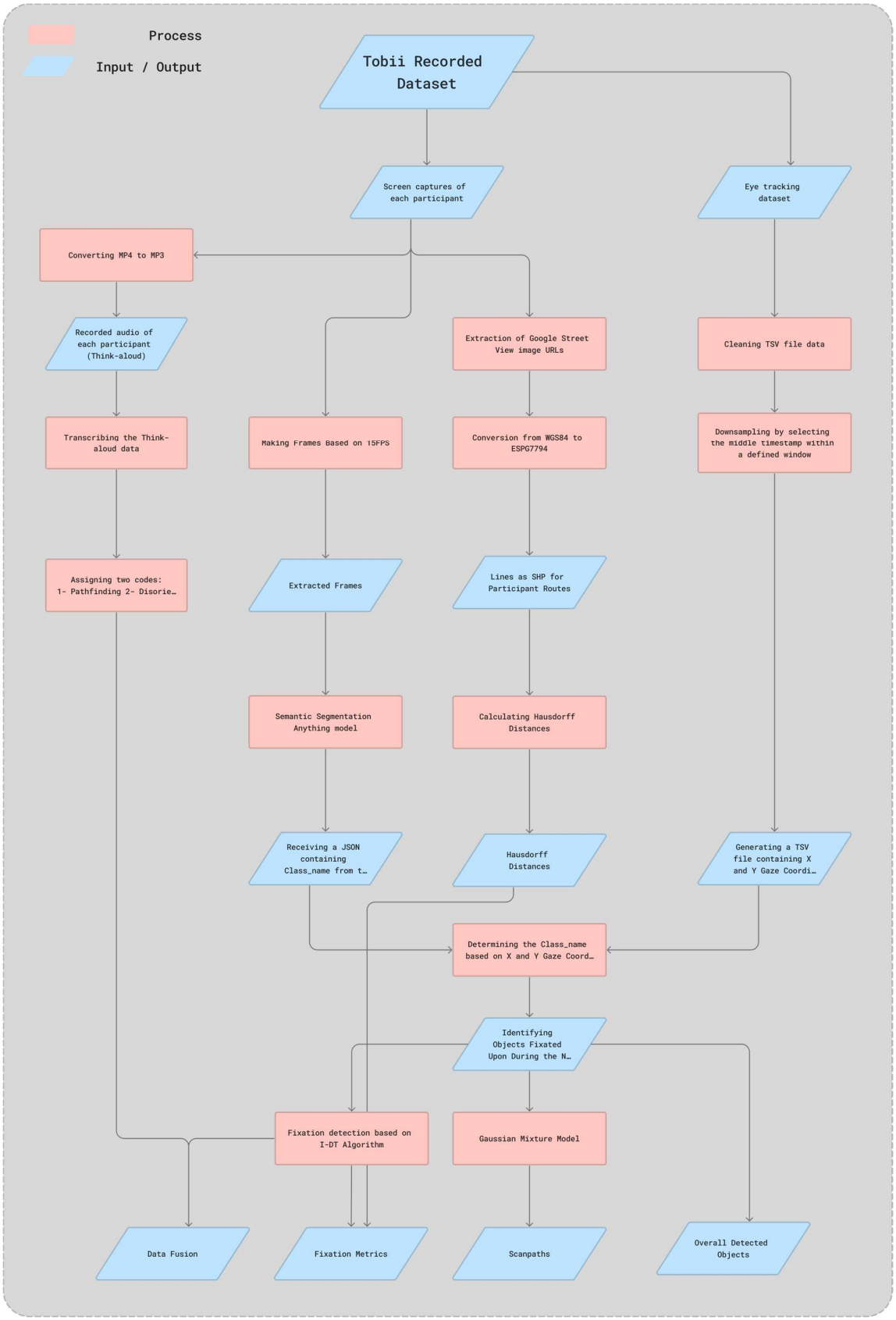


Figure 11 Detailed Methodology

3.4.1. Eye-Tracking Gaze Coordination Cleaning and Downsampling

Following the completion of data collection and acquisition of participant data, the Tobii Pro Lab software generated output in the form of tab separated value (TSV) files. These files contained various metrics, among which the most crucial were the "Recording timestamp [ms]", "Event", "Gaze Point X (in pixels)", and "Gaze Point Y (in pixels)". The data compilation began from the initiation of the recording mode, before calibration, and ended upon termination of the recording by pressing the "Esc" button. Observing the participants' videos, two events were defined, namely "Start_survey" and "End_survey". The "Start_survey" event marked the point where the participant passed calibration and viewed the first frame of the survey, while "End_survey" marked the moment the participant recognized the third landmark or admitted inability to find it, indicating task failure. These specific start and end points of the survey can be identified in the Event column. The recording timestamp [ms] indicates the timestamp of the survey in milliseconds, while Gaze Point X [px] and Gaze Point Y [px] denote the eye position in pixels relative to a 1920x1080 screen.

A challenge was encountered due to the difference in data frequency between the datasets. The eye-tracking sensor captured data at a high frequency of 120Hz, in contrast to the 15Hz of the other videos, creating a mismatch that made direct analysis impractical. To reconcile this discrepancy, a downsampling process was employed, defined as the reduction of samples or data points in a digital signal by discarding or merging existing samples (Boada et al., 2001). Although this process resulted in data loss from the higher frequency dataset, it was essential to make the analysis manageable. The synchronization of frequencies ensured that the data could be accurately compared and analyzed together, streamlining the approach by mitigating the complexity of managing disparate data rates.

In the process of downsampling, it was vital to preserve the original gaze coordinates, represented by each X and Y coordinate, which were determined by the participant's cognitive judgments. Though downsampling inherently involves a reduction in data resolution and can lead to information loss, the selected method ensured that no new gaze points were calculated. Instead, all original gaze points remained intact in relation to the timestamp. This approach maintained the authenticity of the data while aligning with the necessary downsampling requirements.

Moreover, in eye-tracking datasets, temporal occurrence and time of gazes are significant factors (Zemblys et al., 2018) making it suitable to consider the timestamp for downsampling. Accordingly, a Python code was written to execute the downsampling process. The following steps were undertaken in the Python code:

Dividing into Windows: The code breaks the original timestamps into smaller groups called "**windows**," each containing a set of timestamps (**timestamps**) and their corresponding gaze point positions (**x and y**).

Finding the Middle: For each window, the code identifies the middle timestamp and selects the gaze point position recorded at that time (**x_window.iloc**[middle_index] and **y_window.iloc**[middle_index]). This method preserves the central gaze data within the window, considered essential for reflecting the participant's focus during that specific timeframe.

Creating the Downsampled Dataset: By repeating the process for different windows, the code collects the selected gaze point positions and timestamps to create a new, smaller dataset with fewer data points (**downsampled_x**, **downsampled_y**, **downsampled_timestamps**).

Saving the Downsampled Data: The downsized dataset is stored in a new file specified by **output_file_path**. This new file contains the downsampled gaze point positions and timestamps (**downsampled_df**).

In the processing of the eye-tracking data, a downsampling method was employed that selects the middle timestamp within each time window, as shown in Figure 12. This approach offers several notable advantages. By selecting the middle timestamp, the method ensures that the data reflects a balanced and central point within each window, thereby reducing the risk of time skew. Time skew refers to a misalignment or distortion in time series data that could arise if the first or last timestamps were selected, as these might not accurately represent the entire window's activity. The middle timestamp can be seen as more representative since it captures a point equidistant from the beginning and end of the window, potentially encompassing the central tendencies of eye movement within that period. This is especially valuable in eye-tracking data, where variations in event durations and gaze behavior may occur.

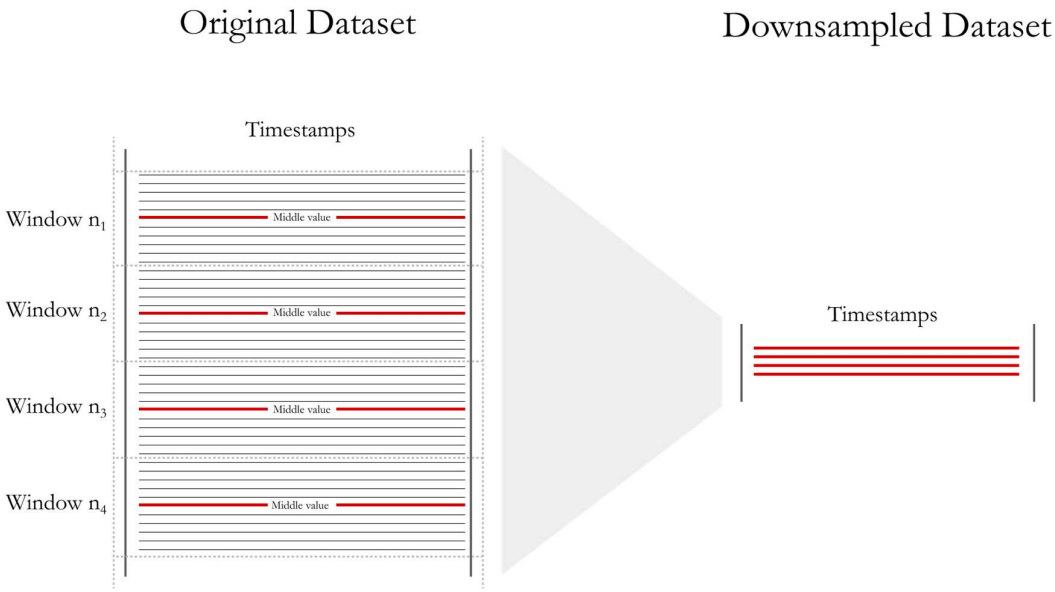


Figure 12 Transformation of Original Data to Downsampled Data

Secondly, the preservation of the original X and Y gaze coordinates is allowed. Instead of introducing calculated or averaged values that might not truly represent recorded gaze points, the integrity of the original data set is maintained.

Finally, for regularly sampled data, the selection of the middle timestamp serves effectively as the selection of the median (Hagler et al., 2011). Given that the median is a robust measure of central tendency, this method ensures the selection of a data point that is less sensitive to outliers or extreme values.

However, this method is not without potential drawbacks. Specifically, a window in this context consists of 9 timestamps, and these 9 are reduced to 1 in the downsampling process, each window lasting about 67ms, corresponding to a frequency of 15Hz. This reduction might result in missing some nuances in the data. One significant consideration is that the middle timestamp might not always capture the most representative or critical data point within each window. Particularly in instances where there is a significant gaze event occurring at the beginning or end of the time window, the middle timestamp may fail to represent this event. This can be seen as a trade-off between reducing noise and maintaining critical information, given the substantial compression of the data from 9 points to just 1. Therefore, while the chosen downsampling method offers numerous advantages specific to the dataset and research aims, it is crucial to remember that the suitability of downsampling methods can vary greatly depending on the specific characteristics of the data and the research question at hand.

3.4.2. Semantic Segmentation

For this investigation, the Semantic Segment Anything tool was utilized to analyze the areas of interest (AOIs) within the navigational survey (Kirillov et al., 2023). This methodology was chosen due to the changing nature of the AOIs throughout the task. The Segment Anything (SA)[1] project is a new task, model, and dataset for image segmentation. Image segmentation involves dividing an image into multiple segments or "masks," where each mask corresponds to a specific object or part of the image. In the context of the SA project, a mask is essentially a labeled area that identifies a particular segment of the image, such as a specific object, feature, or texture. Using an efficient model in a data collection loop, the SA project built the largest segmentation dataset to date, with over 1 billion masks on 11 million licensed and privacy-respecting images. The model is designed and trained to be promptable, so it can transfer zero-shot to new image distributions and tasks. The concept of masks in segmentation allows for detailed analysis and manipulation of individual components within images, and it is central to various applications such as object detection, recognition, and computer vision tasks (Kirillov et al., 2023).

The Segment Anything Model (SAM)[2] presents a robust approach towards random object segmentation, yet its inability to anticipate semantic categories for each mask poses a challenge. The Semantic Segment Anything (SSA) initiative seeks to overcome this drawback by introducing a pipeline that operates in conjunction with SAM to predict the semantic category of each mask. Additionally, SSA embodies an

[1] facebookresearch/segment-anything: The repository provides code for running inference with the SegmentAnything Model (SAM), links for downloading the trained model checkpoints, and example notebooks that show how to use the model. (github.com)
[2] fudan-zvg/Semantic-Segment-Anything: Automated dense category annotation engine that serves as the initial semantic labeling for the Segment Anything dataset (SA-1B). (github.com)

automated annotation mechanism, termed as Semantic Segment Anything Labeling Engine (SSA-engine), which offers comprehensive semantic category annotations for SA-1B or any other datasets, thereby diminishing the need for manual annotation and its associated expenses considerably. (Chen, Jiaqi et al., 2023)

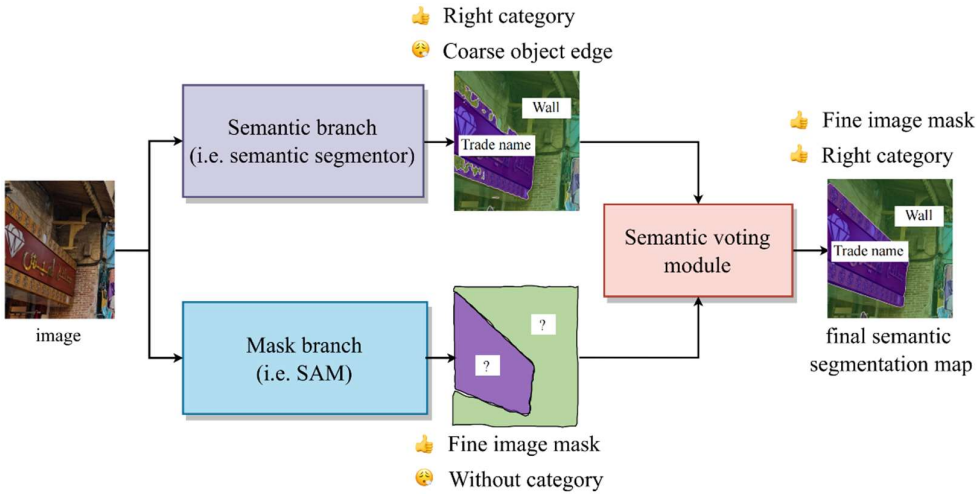


Figure 13 Schematic representation of the Semantic Segment Anything model (Chen, Jiaqi et al., 2023).

The application of a complex model to a large dataset is found to require significant computational resources and time. A challenge encountered relates to the 25 frames-per-second (FPS) videos produced by Tobii Pro Lab. With the duration of these videos, an excess of 183,500 frames are generated, an amount too large for available computational capabilities to process. Nevertheless, the study environment, namely Google Street View, is comprised of unchanged, 360-degree images. This stability suggests that lessening the frame rate should minimally impact the loss of key temporal information during the process of semantic segmentation. During the navigational tasks conducted in Google Street View, participants are primarily engaged in exploring the environment, identifying landmarks, and navigating through the scenes. The emphasis is placed on understanding the spatial layout and recognizing objects of interest, rather than tracking fast movements or temporal changes (Dodsworth et al., 2020). Therefore, a slightly reduced frame rate is expected to retain sufficient information for the semantic segmentation task, while still accommodating the participants' navigational interactions effectively.

The SSA model's output includes PNG and JSON formats. In the PNG format, visually annotated masks are presented (see Figure 14). The JSON file's structure, as shown in Figure 14, includes various sections such as class name, class proposals, segmentation, area, bounding box, predicted IOU, point coordinates, stability score, crop box, and size.

Figure 14 Semantic Segment Anything output in PNG and JSON format

"Segmentation" and "Class Name" are two crucial sections pertinent to this research. They represent the segmentation mask presentation and the classification predicted by the model, correspondingly. In addition, the decoding of the segmentation mask is facilitated with the assistance of Detectron2, an advanced library from Facebook AI Research[3]. (Yuxin Wu & Alexander Kirillov, 2019)

3.4.3. Route Taken by Each Participant

In this research, an analysis is conducted on the path followed by the participants from the starting point to the end point. The exploration of this route can provide insights into the participants' decision-making process regarding route selection, among other factors. To comprehend the path each participant has traversed, a review of their video is necessary, specifically observing the URL of each individual point in Google Street View to ascertain the geographical coordinates utilized.

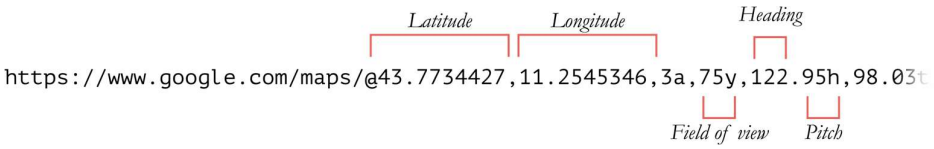


Figure 15 Elements of Google Street View URL

Essentially, each Google Street View URL contains different parameters that can be decoded, as shown in Figure 15. Specifically, the first two parameters of every Google Street View URL correspond to the latitude and longitude of the coordinate system. An extraction of all the points into a TSV file, followed by executing a straightforward python code snippet, allows for the retrieval of the coordinates of all chosen points in the navigation survey for each participant. Subsequently, these points are transformed from the WGS84 coordinate system to EPSG:7794, the latter being specific to Florence, Italy, thereby ensuring accurate distance calculations. From these coordinates, the trajectory each participant followed is constructed and preserved in a shapefile format (SHP).

[3] facebookresearch/detectron2: Detectron2 is a platform for object detection, segmentation and other visual recognition tasks. (github.com)

3.4.4. Think-Aloud Analysis

The initial stage in the think-aloud analysis involved altering the format of participant videos from MP4 to MP3. This was achieved through the use of the "MoviePy" library in Python, resulting in the creation of MP3 files with a bitrate of 192 kbps. Once this conversion process was complete, transcriptions of all the audio files were produced using the "Transcribe" feature in "Microsoft Word," which receives the MP3 file and automatically generates a transcription. Each transcription was then meticulously checked by hand to ensure its accuracy and to rectify any errors that might have occurred. The coded elements of the think-aloud data focused on two primary categories: Pathfinding (I) and Disorientation (II), with statements not relevant to these categories classified as N/A.

Pathfinding refers to the cognitive process participants employ to determine the best route or path to reach their intended destination within Google Street View. It involves making decisions based on visual cues and information provided by the virtual environment. Participants might discuss the selection of specific routes, mention street names they intend to follow, or comment on intersections and turns they are considering. This code allows researchers to analyze participants' navigation strategies, understand the factors influencing their route choices, and identify common patterns in their decision-making during virtual navigation tasks.

Disorientation occurs when participants feel unsure about their position, direction, or how to proceed while navigating in Google Street View. It is characterized by a sense of being lost or confused, resulting in a lack of confidence in making navigation decisions. Participants may voice their uncertainty by stating that they are "turned around," "don't know where they are," or "can't find their way." They might express frustration or use phrases suggesting they feel "lost" or "confused." By applying the "Disorientation" code, researchers can identify specific moments where participants encounter obstacles in their navigation, which helps in understanding the challenges users face while exploring virtual urban environments.

The transcriptions were manually coded, with each statement assigned a respective category or labelled as N/A in correspondence with its timestamp.

3.5. Data analysis

3.5.1. Hausdorff Distances

To gain insights into the navigation choices made by each participant, deviation from an optimal route was calculated for each participant's path. In navigation research, spatial analysis like this is useful as it allows for the observation of how individuals navigate through a given environment.The optimal route in this study was defined as the shortest distance between the start and end points, as traveled along Florence streets through the available Google Street View photosphere locations, not considering the straight-line shortest distance, as shown in Figure 16. Examining deviations from this route can reveal patterns in navigation behavior and provide valuable context for understanding how participants interact with space.

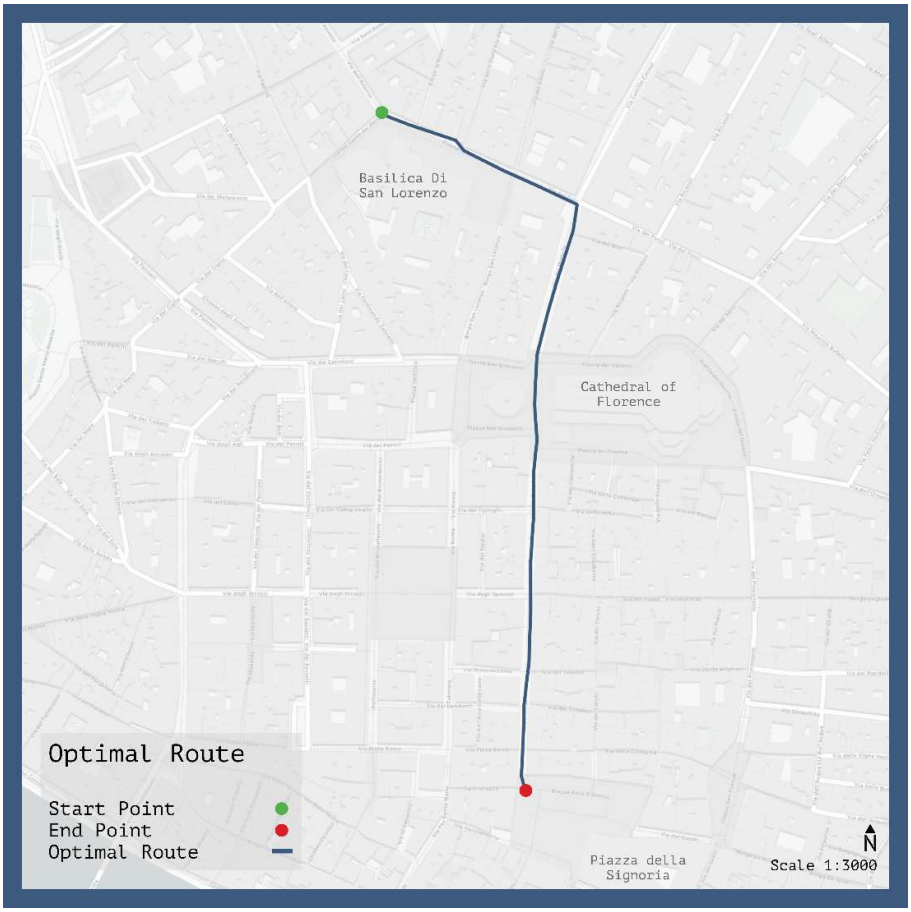


Figure 16 Optimal route on the map of Florence

The Hausdorff distances were employed for this analysis. The Hausdorff distance, named after German mathematician Felix Hausdorff, is a mathematical technique applied to evaluate the likeness or disparity between two sets of points within a metric space. The concept frequently finds use in computer vision, image processing, and pattern recognition, serving to compare two point sets that often signify shapes or contours.

Two point sets, denoted as A and B, have their Hausdorff distance, symbolized as H(A, B), defined by the equation:

$$H(A,B) = \max\left(d(a,B), d(b,A)\right)$$

Here, d(a, B) refers to the least distance from point 'a' in set A to its nearest point in set B, while d(b, A) represents the least distance from point 'b' in set B to its nearest point in set A. Essentially, the Hausdorff distance is determined by finding the closest point in the other set for each point in one set, and then taking the maximum of these distances over all points in both sets. This distance provides a measure of the greatest extent to which the two sets differ.

In fields like computer vision, image registration, shape matching, and object recognition, the Hausdorff distance has various applications. It is employed to measure the similarity or disparity between shapes and proves particularly beneficial when there is a need to compare objects or contours that might have undergone transformations such as translation, rotation, or scaling. By measuring the Hausdorff distance, the alignment of two shapes can be assessed effectively (Birsan & Tiba, 2006).

In Python, the routes taken by participants were processed by loading the corresponding shapefiles (including both participants' paths and the optimal route, with a .shp extension). The following algorithmic steps were implemented to perform the required operations:

Loading Data: The code imports necessary libraries for working with geospatial data, numerical operations, distance calculation, plotting, map visualization, and handling file paths.
It sets up a directory (**output_dir**) to save the output plots.
The optimal route is read from a shapefile and stored in the **optimal_route** variable.
The paths to participant files (representing participant routes) are stored in the **participant_files** list.

Hausdorff Distances Calculation: The code calculates the Hausdorff distances between each participant's route and the optimal route in both forward and backward directions. (i.e., from A to B and from B to A.)
It uses the **directed_hausdorff** function from **scipy.spatial.distance** to calculate these distances.
The calculated Hausdorff distances for each participant are stored in three lists: **hausdorff_distances**, **hausdorff_distances_forward**, and **hausdorff_distances_backward**.

Loop 1 (Calculate Maximum Map Extent): The code iterates through each participant's file to calculate the maximum map extent that includes all the participant routes.
It uses **GeoPandas** to read the participant's file and calculates its extent.
The maximum map extent (**max_extent**) is updated with the bounding box of all routes.
Loop 2 (Plotting and Saving Maps): The code iterates through each participant's file to plot their route, optimal route, and directional arrows to visualize the routes. It uses Contextily to add OpenStreetMap as the background map.
The plot is customized with different markers and colors for **start/end points** and directions. The plot is saved as an image in the output_dir directory, with the participant's name and Hausdorff distance in the filename.

Statistical Calculations: The code calculates the **mean** and **standard deviation** of the Hausdorff distances across all participants. It computes the mean and standard deviation separately for the forward and backward directions.
The results are printed to the console.

3.5.2. Semantic Eye-Tracking Object Detection

3.5.2.1. Objects and fixation

After aligning the semantic segmentation outputs with the downsampled eye tracking data, objects were identified based on gaze coordinates within the same frame. This alignment ensured an equal amount of data in both the eye tracking and semantic segmentation sets, with each data row containing X and Y gaze coordinates present in the segmentation masks for a single point in time.

In Figure 17, a single point at the coordinates (761, 668) symbolizes a gaze location of the participant. When this point is overlaid on the underlying semantic segmentation PNG, it intersects with a mask class as "building". Thus, the code is designed to identify the corresponding class name for each gaze coordinate, indicating what the participant was focusing on. In the gaze dataset, certain points are found to be missing coordinates, arising due to various factors. This could occur when a participant blinks, causing the sensor to fail in detecting the eyes at those specific timestamps, or when the participant looks away from the screen. These instances are accounted for in the dataset, with the missing values denoted as 'N/A'.

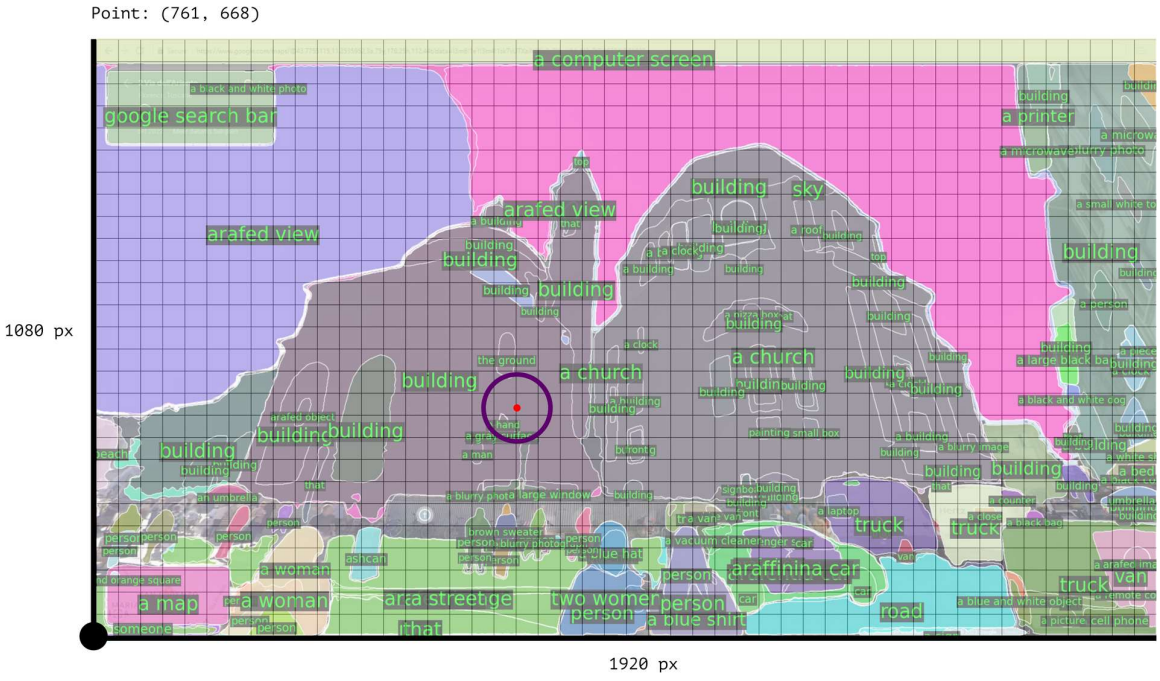


Figure 17 Finding the Gaze coordinates in Semantic Segmentation masks. In this example, the participant is looking at an instance of "building".

Additional metrics, including fixation, were computed. The Identification by Dispersion-Threshold (I-DT) algorithm, as outlined by Salvucci & Goldberg, (2000), was used to detect fixation. This eye-tracking algorithm is employed to differentiate between fixations and saccades in eye-tracking experiments. It requires setting a dispersion threshold and a duration threshold. When the eye-screen distance is known, the dispersion threshold is usually between 1/2° to 1° of visual angle. If this distance is unknown,

exploratory data analysis can estimate the dispersion threshold. The duration threshold commonly ranges between 100 and 200 ms.

Taking into account the stated 1° visual angle and the 1920x1080 resolution of the computer screen, and considering the participants maintained a 60cm distance from the screen, fixation identification based on I-DT demanded that X and Y be within a 34-pixel radius on the screen for at least 200ms. Given the 15Hz frequency of our dataset, where each row corresponds to 67ms, any three consecutive X and Y coordinates within a 34-pixel radius of each other were considered as a fixation.

In Figure 18, a segment of a participant's eye-tracking data is depicted on a 1920x1080 resolution screen. The illustration captures a series of gaze points, connected by lines to indicate their chronological sequence. Gaze points falling within a 34-pixel radius of one another are highlighted in green, signaling a fixation event. Interestingly, not all closely situated points are marked in green, underscoring the significance of time sequence in the identification of fixations.

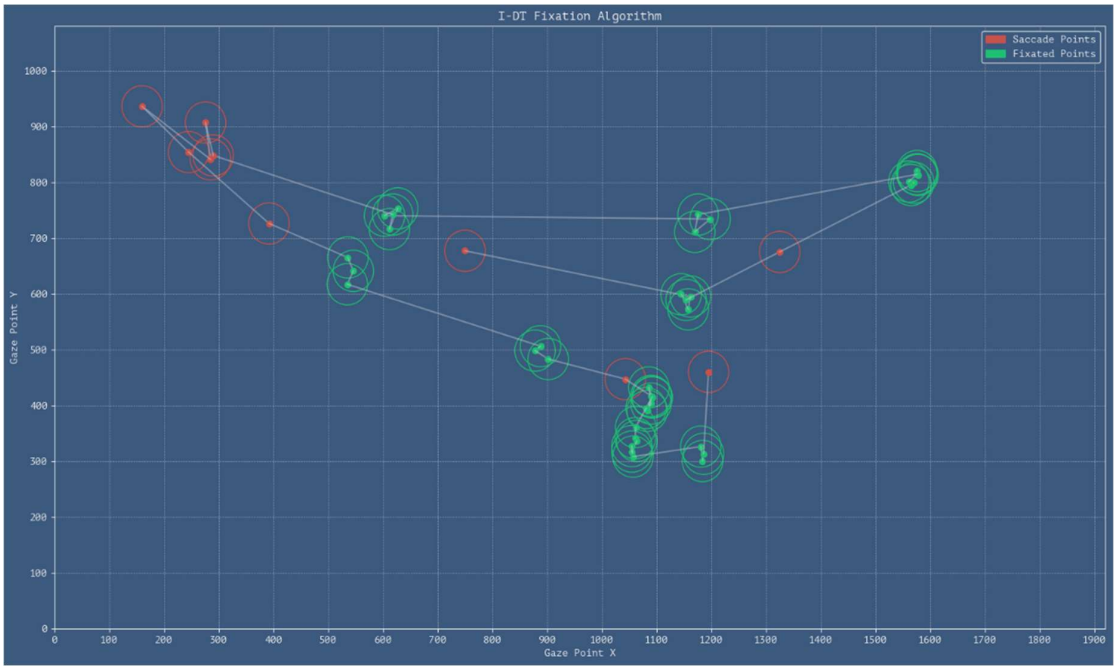


Figure 18 I-DT Fixation Algorithm. Green circles indicate the occurrence of fixations, while red circles indicate the occurrence of saccades.

The following outlines the algorithmic steps implemented in Python to locate the gaze point coordinates within the semantic segmentation JSON files, and to calculate fixation. These steps describe the overall methodology and data processing workflow rather than specific code snippets:

Read Eye Gaze Data: Read eye gaze data from a TSV file named **'Participant_Gaze.tsv'** into a pandas DataFrame called **df_tsv**.

Process Image Segmentation Data: Iterate through JSON files in the directory. Extract the frame number from the filenames using a regular expression pattern and retrieve the **corresponding x** and **y** coordinates of the eye gaze from **df_tsv**.

Find Fixated Objects: For each gaze point (x, y), check if it falls within any of the segmented objects in the corresponding frame from the JSON file. If a match is found, record the class name of the fixated object.

Check Proximity of Gaze Points: Calculate the distance between each gaze point and its neighboring points (previous and next points) using the Euclidean distance formula. Set a **'34 radius proximity'** flag to 1 if the distance is less than or equal to 34 pixels; otherwise, set it to 0.

Identify Fixations: Identify fixations as consecutive gaze points that have a '**34 radius proximity'** flag set to 1 for at least three consecutive points.
Mark these fixations with '**YES**' in the '**Fixation**' column of the DataFrame.

Calculate Fixation Duration and Fixated Object: Set the fixation duration to a constant value of 67 milliseconds for each identified fixation. Determine the dominant class name of the fixated object for each fixation and add this information to the DataFrame.

Save Results to a TSV File: Save the analyzed results, including frame number, filename, class name of the fixated object, fixation flag, fixation duration, and fixated object, to a new TSV file named '**Participant_ObjectDetected.tsv**'.

3.5.2.2. Annotation Filtering and Correction

Upon obtaining the list of objects from the previous analysis, an issue surfaced relating to the annotations that the semantic segmentation assigns to each object. The model's prediction for each object showed significant variation, leading to an excess of unique labels. For instance, labels such as "a tall building", "a building", "building", and "the building" fundamentally denote the same concept, i.e., "building". However, the model generated these as separate, unique labels. This problem demanded a solution, but addressing it was challenging due to the unpredictable variety of generated annotations. The issue couldn't be resolved merely by removing characters or defining dictionaries, given the vast number of frames.

After exploring various methods, including techniques to find common characters among labels and defining dictionaries to find and group words, the Natural Language Toolkit (NLTK) was found to be most effective for the task. NLTK is a collection of libraries and programs designed to perform symbolic and statistical natural language processing (NLP) tasks, specifically for the English language, using Python. It provides a variety of functionalities, such as classification, tokenization, stemming, tagging, parsing, and semantic reasoning, that were more suited to the requirements of the project (Bird et al., 2009).

Applying NLTK to the problem proved to be highly effective. It allowed the filtering and cleaning of annotations in different stages, with each stage eliminating a portion of the undesired data. This method

outperformed the previous techniques used, such as finding common characters among labels and defining dictionaries to find and group words. The subsequent section details the algorithmic steps performed on the labels, describing the procedures used in the filtering and cleaning process.

Importing Libraries: nltk, pandas, and specific modules from nltk (stopwords, word_tokenize, pos_tag, and WordNetLemmatizer).

Downloading NLTK Resources: Such as the tokenizer, averaged perceptron tagger, WordNet, and stopwords. These resources enable specific text processing tasks later in the code.

Text Preprocessing Function: The code defines a function **preprocess_label(label)** that standardizes the textual labels in the dataset, ensuring that equivalent terms are represented in a consistent manner. The **preprocess_label(label)** function takes a label as input and processes it through the following steps:

Handling Missing Labels: If the label is "N/A" (i.e., missing), it returns the label unchanged.
Lowercasing: It converts the label to lowercase.
Tokenization: It breaks the label into individual words.
Stopword Removal: It removes common English stopwords from the tokenized words, except for the word 'it.' (This word is retained due to its high frequency and relevance to the model.)
Part-of-Speech Tagging: It assigns a part-of-speech tag to each remaining word.
Lemmatization: It converts the words to their base or dictionary forms, focusing on nouns (tagged as 'NN') or the word "it." This step contracts essentially equivalent terms down to single, standardized versions.
Loading the Dataset: The code loads a dataset from a TSV (tab-separated values) file, containing columns named **"Class Name"** and **"Fixated_object."**

Applying Text Preprocessing: The **preprocess_label** function is applied to the "Class Name" column and the "Fixated_object" column in the DataFrame df, ensuring that the data is represented in a consistent and standardized form.

Tokenization, in the realm of natural language processing (NLP), signifies the segmentation of a text into individual units or tokens, usually words or subwords (Rai & Borah, 2021). Stopwords, often excluded from web searches for quicker indexing and parsing, are frequent words that tend to be filtered out during or post natural language data processing due to their lack of significance (Rajaraman & Ullman, 2011).

Part-of-speech tagging categorizes words into different parts of speech, such as adjectives, adverbs, nouns, and verbs, according to their semantic roles and syntactic functions in a sentence. This tagging is used here to identify and remove adjectives from labels.

Lemmatization is a technique in NLP that groups inflected forms of a word into a single base form or lemma with a unified meaning (Müller et al., 2015). For instance, “running” is lemmatized to “run,” and “cars” to “car.”

Even after employing the Natural Language Toolkit (NLTK), some issues in the labels remain due to inherent complexities:

1. Model-related issues: The predictive model sometimes generated labels that were nonsensical and didn't pertain to any identifiable object or concept. Examples include redundant repetition of words or symbols, such as "a blurry blur effect effect effect effect effect..." or "metal vent vent vent vent vent vent...". There were also instances where unrelated symbols were combined to form nonsensical labels like "thermo®®™TMTMTMTMTMTMTMTM."
2. Environment-specific issues: Some labels were a product of the task environment itself. For instance, the labels like "blurry screen" or "blurry effect" were generated during the transitions between images in Google Street View. When the next image was loading and no clear objects could be detected, these placeholder annotations were created. Additionally, the predictive model created a multitude of labels concerning the screen when participants looked towards the edges of the screen or parts of the browser that were visible during the video recording of their performance.
3. Inadequate NLTK processing: Certain labels did not get effectively processed during the NLTK cleaning stage due to structural characteristics. For example, annotations using hyphenation, such as "wall-stone," posed a challenge to the cleaning process.
4. The need for semantic grouping: There were labels that, while accurately segmented and cleaned by the NLTK, could be semantically grouped together for more meaningful analysis. For instance, the labels "street", "pavement", "sidewalk", and "road" were semantically related and could be grouped under a larger category, "thoroughfare". This regrouping aimed to provide a better, consolidated understanding of the types of objects identified by the semantic segmentation.
5. Labels with ambiguous semantic meaning: Some labels assigned by the semantic segmentation didn't carry clear semantic meaning, posing challenges for analysis. Labels like "it" and "top" are examples of this category. While these labels might make sense in specific contexts, they lacked overall clarity and failed to provide meaningful insights when analyzed in isolation.

Label merging was conducted manually, aided by Python code, which tracked and recorded the changes as text files. Each transformation created its own text file, enabling overall change tracking across different labels. In Figure 19, one of the merged label groups is depicted. The figure illustrates how multiple labels related to vehicles, including brand-specific labels like "audi car" or "tesla car," as well as service vehicles such as "police van," have been consolidated into a single label called "vehicle." All label changes are detailed in Annex 2.

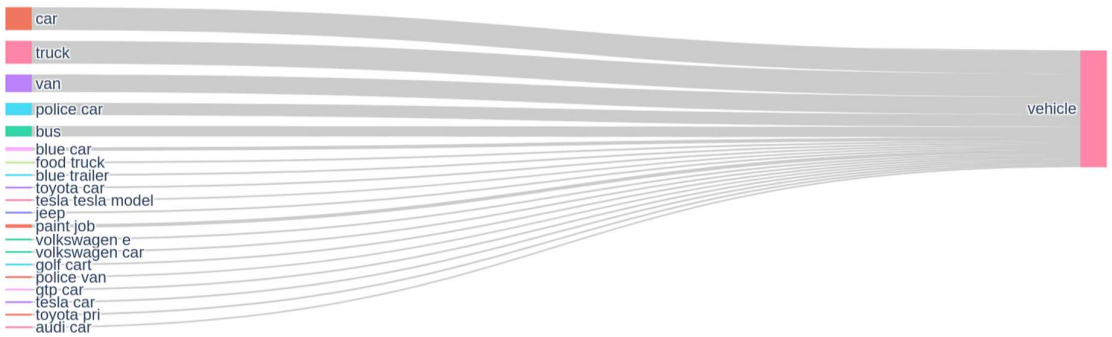


Figure 19 Example of Label Transformation

3.5.2.3. Heatmap and Scanpath Creation

Building on the fixation and saccade metrics computed in section 3.5.2.1 , another analysis related to the overall fixation points across the screen during the entirety of the navigation task was carried out. This study, known as a scanpath analysis in eye-tracking research, plots the pattern of fixations and saccades generated by eye movements over a certain period (Holmqvist et al., 2011). Scanpaths, which represent the eye movement patterns induced by specific tasks, provide deeper insights into intricate strategies by depicting temporal, spatial, or both types of characteristics in more detail (Byrne et al., 2023).

While the Tobii Pro Lab software does offer scanpath analyses, due to the downsampling of the dataset, the scanpaths were re-plotted using the downscaled data through a customized implementation. The I-DT algorithm, as previously described, was applied for detecting fixations. Two distinct analyses were conducted on fixations and scanpaths, utilizing techniques adapted to suit the specific requirements of the study.

Scanpaths exhibit the distribution of overall fixations on the screen, connected by saccade lines. Faint grey lines were employed to link fixations across the screen. As fixation durations increase, the radius of the fixation circle slightly expands. However, when considering the entire duration of the navigation task in this plot, the incremental increase in the radius for each fixation is quite small, resulting in subtle variations in circle size.

```
Data Processing: The eye-tracking data from TSV files for multiple participants
are read and processed. It calculates the Euclidean distance between consecutive
gaze points to identify fixation points (when the distance is within a certain
threshold).
Fixation durations are determined based on consecutive fixation frames.

Data Scaling: Fixation durations are scaled to a range of [0, 1] using
MinMaxScaler, a preprocessing method that transforms numerical values into a
specific range by using the minimum and maximum values of the feature. This
scaling helps to normalize the fixation durations, making them suitable for
further analysis or visualization.

Data Visualization: The code uses matplotlib to create a scatter plot for each
participant. Each fixation point is represented as a small dot with a color
corresponding to its assigned cluster. The cluster centers are plotted as '+'
symbols with circles overlaying them. A legend is added to the plot, indicating
the cluster number and the number of points in each cluster.

Loop Through Participants: The code processes each participant's data file one by
one in a specified directory. It extracts participant numbers from the filenames
and incorporates them into the plot titles.
```

The second form of analysis employed on the fixation data was the generation of heat maps. This technique enables the visual inspection of the spatial distribution of eye movements across a given field. By using this, the patterns of eye movements can be revealed for individual participants or for multiple ones. Rather than capturing the sequence of fixations, heat maps focus on the spatial distribution of these fixations (Grindinger et al., 2011). These maps typically employ Gaussian Mixture Models (GMMs) to

denote the frequency of fixation points. The color intensity of the maps corresponds to the quantity of fixations at each point, allowing for effective comparison of areas with varying levels of visual attention (Hill et al., 2020).

For this analysis, GMMs were initially utilized to construct heat maps of fixation data for both successful and unsuccessful participants. GMMs are probabilistic models that assume data points are produced from a finite number of Gaussian distributions with unknown parameters. Using the sklearn.mixture.GaussianMixture class from the sklearn library, the x and y coordinates of fixation points were input to generate a heat map. Each point on this map signifies the estimated density of fixation points at that location.

After the generation of these heat maps, a process of thresholding was applied to create binary masks, simplifying the heat maps into high ('1') or low ('0') density regions. The optimal threshold value was determined using Otsu's method, an automatic thresholding technique that calculates the threshold by minimizing the intraclass variance of the black and white pixels. This method classifies the pixels into two classes by finding a threshold that minimizes the variance within each class and maximizes the variance between classes, based on the image histogram (Sezgin & Sankur, 2004).

Intersection over Union (IoU) was then calculated between the binary masks of successful and unsuccessful participants. IoU, a metric that ranges from 0 to 1, measures the overlap between two binary masks. The value is calculated as the area of intersection divided by the area of union, providing a quantifiable measure of the similarity between the fixation patterns of different groups.

Finally, the centroid distance was calculated. This metric measures the Euclidean distance between the centers of mass of high-density regions for both groups of participants. The term "mass" in this context refers to the density of fixation points. The overall process executed can be summarized as follows:

Importing Libraries: It imports necessary libraries, including pandas, numpy, matplotlib.pyplot, scipy.ndimage.measurements, skimage.filters, sklearn.mixture.GaussianMixture, and glob.

Process Directory Function: **process_directory**(directory) takes a directory path as input. It finds all TSV files in the directory using the glob function.
For each file, it calls the **process_file** function (defined in the previous code) to extract fixation points.
The fixation points from all files are combined into one array and returned.

Process File Function: This function reads a TSV file, identifies fixation points, and returns a DataFrame of fixation coordinates.

Create Heatmap Function: **create_heatmap**(fixations) takes an array of fixation points as input. It fits a **Gaussian Mixture Model** (GMM) with 3 components to the fixation points. A grid of points is created, and the densities are predicted for each point using the GMM. The negative log-likelihood scores are reshaped into a density grid and returned.

Calculate IoU and Centroid Distance Function:
calculate_iou_and_centroid_distance(successful_mask, failed_mask) takes two binary masks (successful and failed density maps) as input. It computes the Intersection over Union (IoU) between the masks, indicating the similarity of high-density regions. The centroid distance between the centroids of high-density regions in both masks is calculated.

Load Fixation Data: Fixation data is loaded for successful and failed navigation scenarios using **process_directory** for both directories **'Succesful'** and **'Failed'**.

Create Heatmaps: Heatmaps are created for successful and failed navigation scenarios using create_heatmap.

Compute Otsu's Threshold: Otsu's threshold is calculated for the successful and failed density maps using filters.threshold_otsu.

Apply Thresholds and Get Binary Masks: Thresholds are applied to the successful and failed density maps to obtain binary masks using the calculated thresholds.

Calculate IoU and Centroid Distance: IoU and centroid distance are computed for the binary masks representing high-density regions in successful and failed scenarios using **calculate_iou_and_centroid_distance**.

3.5.2.4. Event-based Locomotion and Observation

The task of merging different datasets, specifically the spatial data of a participant's route and eye-tracking data regarding fixated objects, involves careful consideration. Both datasets have synchronicity at their start and end times, but this alone is insufficient. The assumption of a constant speed of movement along the path based on just the start and end times might not reflect the reality. To counteract this issue, two states, "locomotion" and "observation", were introduced in the eye-tracking dataset. These states signify whether the eye-tracking fixation or saccade happened during movement (locomotion) or during a pause (observation). Incorporating these states helps in overlaying the two datasets by accounting for variations in speed and moments of stillness. However, the approach assumes a constant speed during the locomotion state, as the inclusion of variable speeds would necessitate complex calculations.

The process of defining locomotion and observation states involved the examination of participants' videos in the Tobii Pro Lab software. Two events, "StartLooking" and "StopLooking", were marked. When a participant paused to examine 360° images, the "StartLooking" event was activated, and the "StopLooking" event was recorded when the participant chose their route. Any eye-tracking gaze coordinates falling between these two events were marked as observations, and gazes outside these events were designated as locomotions.

Once these states were assigned, each point in the eye-tracking dataset fell under either the locomotion or observation state. By using interpolation with the spatial data, the locations of these points can be determined. The duration of each event also plays a part, with each row of data representing 67ms at a frequency of 15Hz. Thus, the length of each pause can be measured. Interpolation was carried out using Python code, and the following outlines the steps taken within the code.

Load Data: It loads two data files a shapefile (**shp_file_path**) containing a **LineString** representing a participant's path, and a TSV file (**tsv_file_path**) containing additional information about observation and locomotion states for each frame in the path.

Converting LineString to Points: The code converts the **LineString** in the shapefile to a sequence of points. The number of points equals the number of rows in the TSV file.

Preparing the DataFrame: The code adds the points and the frame numbers to the TSV **DataFrame (df)** as new columns: **'geometry'**, **'Marked'**, and **'Frame'**.

Marking Rows for Interpolation: The code determines which rows to mark for interpolation based on state transitions between **'observation'** and **'locomotion'**, as well as setting a fixed number of rows (N) between each **'locomotion'** point.

Interpolating Points: It calculates the number of marked rows and then interpolates points for only the marked rows based on the **LineString** geometry.

Updating the DataFrame with Interpolated Points: The code updates the **'geometry'** column of the DataFrame with the interpolated points for the marked rows.

Creating a GeoDataFrame: A new GeoDataFrame **(gdf_marked)** is created, containing only the marked rows with their interpolated points as the geometry column.

Saving as Shapefile: The GeoDataFrame containing the marked rows and interpolated points is saved as a new shapefile at the specified path (**output_shp_file_path**).

3.5.3. Extraction of Data from Think-Aloud Transcripts

After creating the segmentation codes, each transcript was thoroughly examined, with participant statements allocated respective codes within an Excel spreadsheet. These codes were assigned in response to both explicit and implicit participant statements.

A Python script was employed to analyze the Excel files after coding. The script essentially quantified each assigned code and estimated the duration of the associated statement. The estimation of duration relied on the time difference between a statement and the subsequent one.

However, it's critical to acknowledge certain considerations when interpreting think-aloud data. The duration tied to each code during the think-aloud protocol can be influenced by various factors. These could include participants' hesitation in verbalizing all their thoughts, interruptions in their thought process, and the timing of the next statement. Hence, the computed duration for each code may not always precisely reflect the actual time spent on a specific cognitive process or behavior.

Given these potential limitations, it's essential to interpret the duration data with caution and consider them as indicative rather than definitive measures of the exact time spent on each behavior. The think-aloud protocol provides valuable insights into participants' thought processes and behaviors, but it is not a perfect representation of real-time cognitive activities.

3.6. Ethical Considerations

This research involved human participants, making the inclusion of ethical considerations predominant. In addition to the specifics of this study, general principles such as respect for persons, beneficence, and justice were also adhered to throughout the research process.

Prior to the commencement of the navigation survey, participants were informed about their rights, the potential risks involved in the survey, and how their data would be used. This was done to ensure informed consent, an integral part of ethical research involving human subjects. Each participant signed a consent form, signifying their voluntary participation in the study.

To protect the participants' identities and maintain confidentiality, a process of pseudonymization was employed. Participants were assigned unique identification codes, allowing for data analysis without revealing identifiable personal information. This step was taken to prevent any potential harm that could arise from the inappropriate use of personal data.

Data was stored securely to prevent unauthorized access, and all results were reported honestly and accurately, with no manipulation to fit preconceived assumptions or hypotheses. Finally, at the conclusion of the study, all participants were debriefed and had the opportunity to ask any questions about the research. The ethical commitment extended beyond data collection, maintaining respect for participants' rights and dignity throughout the entire research process

4. DISCUSSION AND RESULTS

This chapter examines the results of the navigational tasks completed by the participants. Table 2 details the duration of the task for each participant, categorizing their performance as success or failure. The table outlines the differences in completion times and outcomes across the participants, offering a clear view of the performance in the navigation tasks. These results may lead to further insights into the elements affecting navigation abilities. The following sections will provide a closer analysis of the data presented in Table 2. This analysis will focus on understanding the specific factors involved in successful and failed navigation attempts, shedding light on the underlying mechanisms. This examination aims to contribute to the ongoing research in navigation, focusing on the dynamics observed in this study.

Number of Participant	Duration of Task	Performance
Participant 01	00:03:57	Succeed
Participant 02	00:14:06	Succeed
Participant 03	00:03:05	Succeed
Participant 04	00:07:19	Succeed
Participant 05	00:03:24	Succeed
Participant 06	00:04:34	Succeed
Participant 07	00:16:18	Succeed
Participant 08	00:02:22	Succeed
Participant 09	00:06:55	Succeed
Participant 10	00:14:29	Failed
Participant 11	00:20:04	Failed
Participant 12	00:05:57	Succeed
Participant 13	00:19:48	Failed

Table 2 Task duration for each participant

4.1. Identification of Objects and Fixations

4.1.1. Identification of Objects

After analyzing the data, the final plots indicating the various objects each participant observed during the navigation task were obtained. The variety of objects in each participant's plot depended on which objects they looked at and the duration of their task. As shown in Table 2, which provides the task duration for each participant, the variety of observed objects increased for participants numbered 02, 07, 10, 11, and 13, whose tasks took longer than those of the other participants.

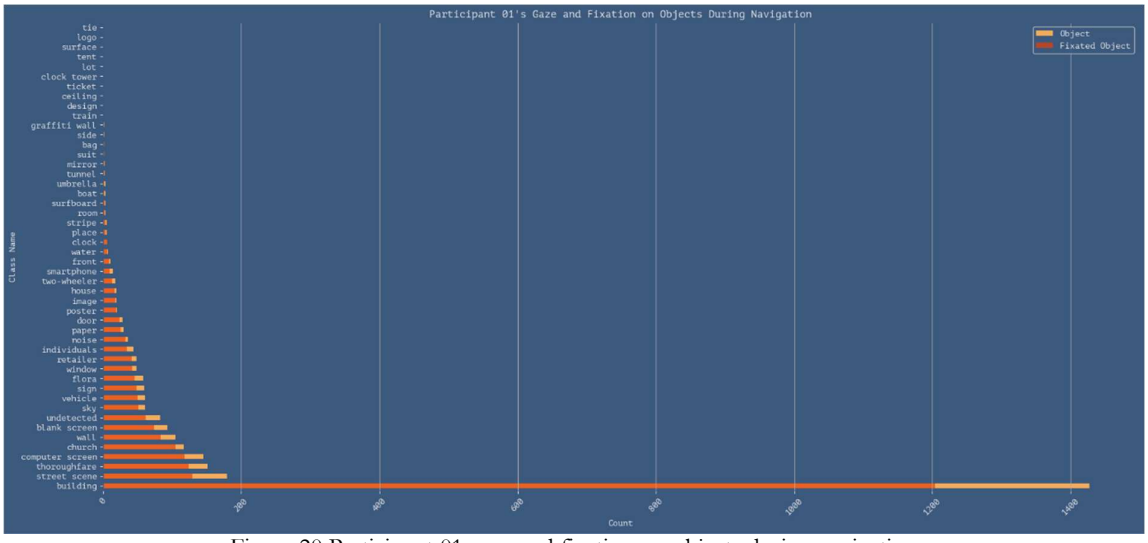


Figure 20 Participant 01 gaze and fixation on objects during navigation

The 'building' object dominated the results across all participants, which is predictable considering the urban scene. The second most observed object varied between participants, although 'thoroughfare' was a common choice for many. However, as show in Figure 20 and Figure 21this pattern was not common across all participants, and participant numbers 01 and 08 did not regard 'thoroughfare' as the second most viewed object. This could be related to the fact that these two participants concluded the task in very short times of 00:03:57 and 00:02:22, respectively. The list of objects for all of the participants can be found in Annex 3.

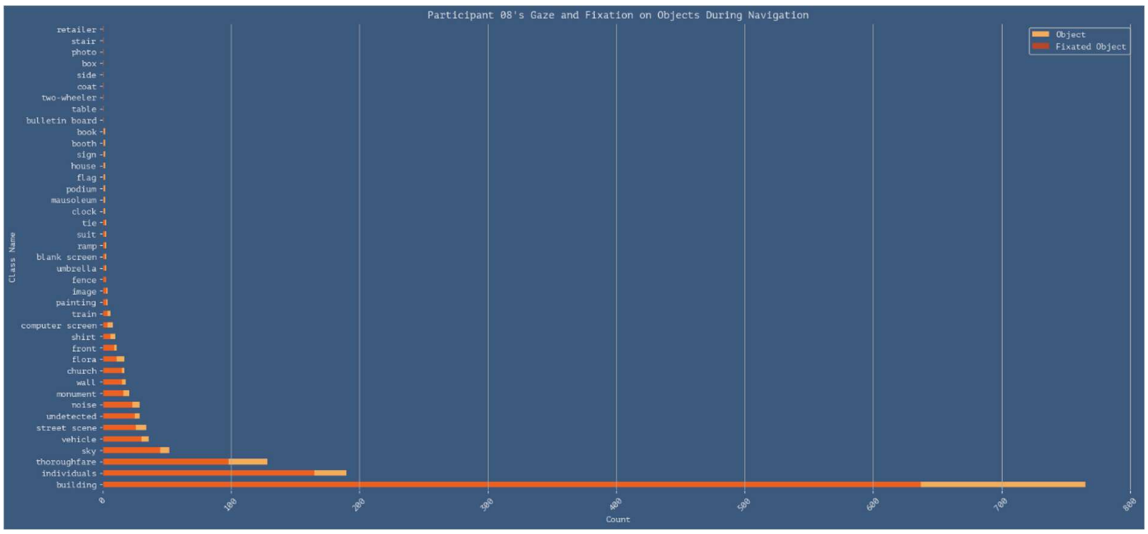


Figure 21 Participant 08 gaze and fixation on objects during navigation

These variations in object observation, fixations, and task performance may indicate different approaches taken by each participant when navigating an urban environment. These differences could be a reflection of individual preferences or understanding of the scene. To understand the participants' strategies, it might be useful to focus on the objects commonly noticed by all participants. Shared points of focus could provide information about common aspects that attract attention. Upon completion of the navigation task analysis, 16 common objects were identified that were observed by all participants.

Upon completion of the navigation task analysis, 16 common objects were identified that were observed by all participants. These include building, vehicle, street scene, suit, thoroughfare, noise, wall, image, sky, undetected, front, room, individuals, blank screen, computer screen, and church.

4.1.1.1. Heatmap of Commonly Identified Objects

During the study, gaze patterns of participants were captured and examined while they performed navigational tasks involving different objects. Distinct disparities in the frequency with which participants observed these objects are depicted in logarithmically scaled heatmaps of gaze counts, using the natural logarithm (base e) as showcased in Figure 22. The reason for using a logarithmic scale with base e is because certain objects, like buildings, have significantly high values. These high values can make it hard to understand the differences between various objects. By using this specific scale, the range of values is made smaller, which makes it easier to notice and compare high and low frequency observations. This helps us better understand the gaze data and what participants were focusing on during navigation.

In the heatmap presented in Figure 22, the first three participants are those who failed the task, and they are followed by the rest of the participants who succeeded. This ordering provides additional context to the variations in gaze patterns, enhancing the insights into how failure or success in the task may be related to the participants' focus on different objects during navigation.

Upon examining the gaze data, it is noted that 'building' consistently held the highest counts among all participants, indicating that buildings were the objects most frequently observed during navigation. The reason for this may be due to 'buildings' large size, noticeable presence, and the critical information they provide about an individual's location and the overall layout of the city.

On the contrary, 'room' and 'front' were found to have the lowest counts. These low counts might suggest these objects were of lesser importance to the participants during their navigation, possibly because of their less distinctive features or their lower relevance for orientation within the city.

Interestingly, a considerable variation in the viewing frequency of 'thoroughfare', 'wall', and 'individuals' was observed among participants. Certain participants (10, 11, 13) displayed high counts for these objects, while others (01, 03, 04, 05) recorded substantially lower counts. This discrepancy might suggest diverse navigation strategies or perceptual tendencies among the participants. Participants who failed in the task (10, 11, 13) showed greater interest in 'thoroughfare', 'wall', and 'individuals.' This could indicate more scattered gaze behavior, suggesting an attempt to gather more information from their environment beyond just focusing on 'building'. It might also be the case that these participants were focusing on less useful things, perhaps objects with less uniqueness or location permanence (e.g., 'thoroughfare' or 'individuals'), which could have hindered their navigation success. This observation could be interpreted as a possible reflection of their navigation strategies, where a broad focus may not necessarily contribute to successful task completion. This provides valuable insight into the relationship between gaze behavior and task performance in navigational contexts.

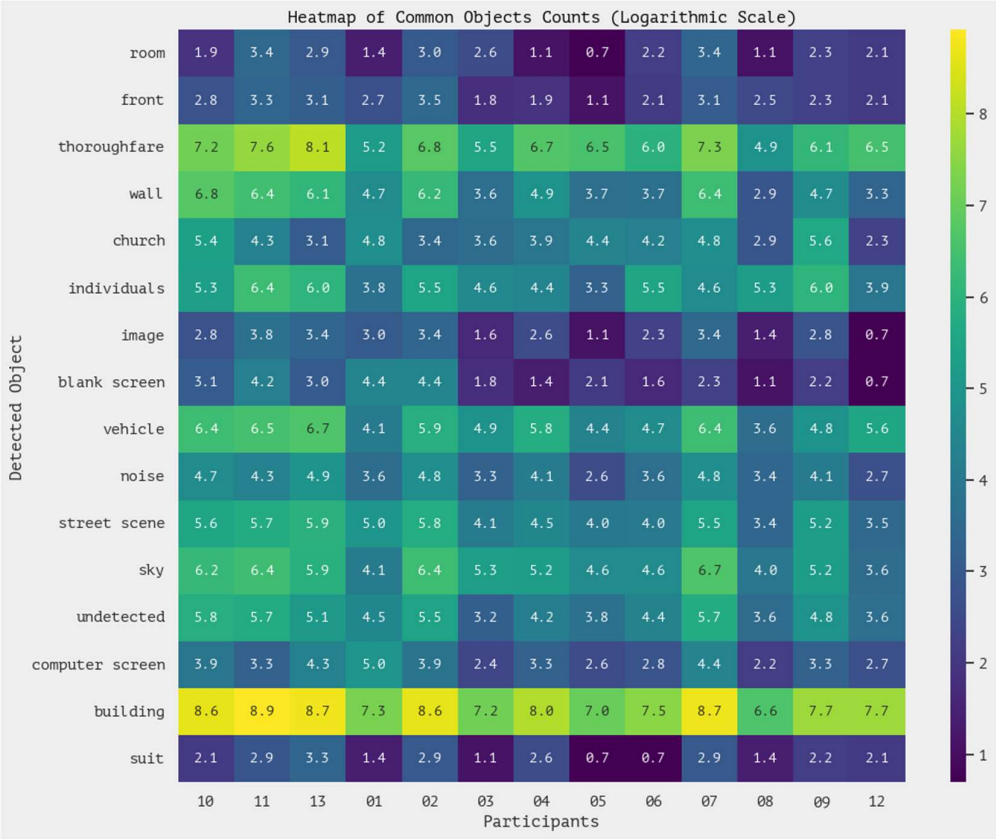


Figure 22 Heatmap of common objects counts (logarithmic scale)

4.1.1.2. Performance Correlation of Identified Objects

In the navigation task, where the goal is to move from a starting point to an endpoint, the time taken plays a critical role in analyzing participant performance. The duration of time for each participant, converted into seconds, is utilized as a 'performance factor.' As shown in Table 2, this factor represents the duration of the task for each participant and is used as a key measure for analysis. The 'performance factor' allows us to calculate correlations between each participant's performance and their gaze upon common objects. This method provides a numerical way to link the performance of participants with their attention to commonly viewed objects in their surroundings, giving a precise quantitative understanding of how gaze behavior relates to navigational success.

These are interpretations based on statistical correlations and do not definitively establish cause-and-effect relationships. They merely suggest trends or patterns that may exist in the data:

Object	*Correlation*
building	0.989
vehicle	0.959
street scene	0.916
suit	0.886
thoroughfare	0.883

noise	0.879
wall	0.871
image	0.869
sky	0.858
undetected	0.853
front	0.826
room	0.799
individuals	0.622
blank screen	0.362
computer screen	0.263
church	0.075

Table 3 Common identified objects correlation table

Building (0.990), vehicle (0.959): A robust positive correlation exists between the performance time and the frequency of looking at buildings and vehicle. The results indicate that the objects participants looked at during the navigational task have varying degrees of correlation with the time taken to complete the task, representing performance. Strong positive correlations with objects such as buildings and vehicles suggest that participants who spent more time looking at or focusing on these objects tended to take longer to complete the task, indicating worse performance. This might imply that there is something about these objects that distracts participants or requires more cognitive processing, leading to longer completion times. Considering that 'building' is the dominant label, focusing on it may not yield particularly insightful results. However, the correlation with 'vehicle' is intriguing, suggesting that gazes on vehicles might play a misleading role in navigation.

Street scene (0.916), suit (0.886), thoroughfare (0.883), noise (0.879), wall (0.871), image (0.869), sky (0.858), undetected (0.853), front (0.826), room (0.799): In these cases, the relative positive correlations suggest that participants who frequently observed certain objects, such as 'street scene' and 'thoroughfare,' may have taken more time to complete the task. These relatively strong correlations imply that these specific objects could play a key role in understanding participant performance. On the other hand, moderate positive correlations with objects like 'front' and 'room' reveal a more nuanced relationship. Although an increased focus on these objects is associated with somewhat longer completion times, the connection is not as clear-cut as with the objects that have stronger correlations. As a result, while these objects may still be relevant to performance, they may not be as crucial or distracting.

Computer screen (0.263)[4] and church (0.075): A positive but weaker correlation with certain objects, such as the computer screen, suggests that participants who often looked at them might have taken longer to finish the task. Other weak positive correlations, such as the one with the church, reveal only a slight tendency for increased focus on these objects to be associated with longer completion times. These objects likely have minimal to no significant impact on performance in the navigational task.

[4] It should be noted that the revised segmentation output of the SSA model has identified the border of the browser on the screen as the 'computer screen'.

These correlations suggest patterns but do not definitively prove causal relationships. For instance, longer task completion times may result from more thorough exploration rather than being hindered by looking at specific objects. Performance is measured by the duration taken, with longer times indicating lower performance, so positive correlations suggest that a higher frequency of viewing these objects is associated with lower performance. Negative correlations, had they been present, would have indicated the opposite. The analysis provides insights into how attention to different objects relates to performance in the navigational task. Objects with strong correlations may be key points of interest or distraction, possibly hindering performance. Meanwhile, objects with moderate or weak correlations offer additional context but may not be as directly influential.

4.1.1.3. Hypothesis Testing

In the context of eye-tracking research, "dwell time" is a term that refers to the total amount of time an observer's gaze remains fixed on a particular area of interest (AOI) within the observed scene (Hofmaenner et al., 2021). This measure is often used to gain insight into what a participant is focusing on during a given task, as longer dwell times typically suggest that an individual is spending more time processing or considering the information in a specific area. Thus, by analyzing the dwell time, we can gain a better understanding of how participants visually interact with and navigate through different urbanscapes.

The choice of statistical test used in the analysis of our data was primarily influenced by the sample size, and the methodological alignment with similar research in the field of eye-tracking. Given that the sample size for this study was relatively small, the Mann-Whitney U test was chosen as an appropriate non-parametric test for the analysis. This test is particularly suitable for small sample sizes and does not make any assumptions about the distribution of the data. Furthermore, its application is well-established in eye-tracking research, as evidenced by previous studies such as those conducted by Caldani et al., (2020) and Cheng et al., (2022). The Mann-Whitney U test provides a robust method for comparing independent samples and assessing whether there is a significant difference in the dwell time of AOIs between participants who successfully completed the navigation task and those who did not. This alignment with accepted practices in the field strengthens the methodological foundation of our analysis.

AOI Object	*Mann-Whitney U Statistic*	*p-value*
Computer Screen	7.5	0.236
Room	8	0.271
Thoroughfare	1	0.014
Front	4	0.076
Individuals	3.5	0.063
Noise	4	0.077
Wall	2	0.028
Vehicle	1	0.014
Street Scene	2	0.028
Church	13	0.8
Sky	5	0.112
Suit	6.5	0.174
Image	5.5	0.128

Blank Screen	6	0.15
Building	2	0.028
Undetected	3	0.04

Table 4 Mann-Whitney U Test Results for Dwell Time

In interpreting the results, the p-value is a critical factor. A p-value less than 0.05 is often considered statistically significant, suggesting that the observed differences are unlikely to have occurred by chance alone.

Here, we can see that the 'thoroughfare', 'wall', 'vehicle', 'street scene', 'building', and 'undetected' AOI objects all have p-values less than 0.05, indicating a significant difference in dwell time for successful and unsuccessful participants for these objects.

The other AOI objects, like 'computer screen', 'room', and 'church', among others, all have p-values greater than 0.05, indicating a lack of statistical significance in the differences observed in dwell times between the successful and unsuccessful participants.

Based on these results, Hypothesis 1, which posits a significant correlation between the dwell time of AOI objects among successful and unsuccessful participants, is partially supported. A significant correlation exists for some AOI objects (i.e., 'thoroughfare', 'wall', 'vehicle', 'street scene', 'building', 'undetected'), but not all. Therefore, while the hypothesis holds true for some AOIs, it is not universally applicable across all AOIs observed in this study.

4.1.2. Fixated Objects

4.1.2.1. Heatmap of Commonly Fixated Objects

A heatmap was also created for fixation counts in this study. This visual representation highlights which objects not only attracted participants' gaze but also sustained their fixation. A fixation often points to a higher level of visual engagement, helping us understand which objects participants deemed significant or engaging.

According to the heatmap, 'building' continued to be the object most often fixated on by participants. This reinforces the important role buildings play in navigating a city. A high fixation count on buildings suggests that participants weren't just glancing at them but taking time to gather detailed information. In contrast, 'front' showed the fewest fixation counts, which aligns with the gaze data and could indicate that these objects were of less interest to the participants. The variation in fixation counts for 'thoroughfare', 'individuals', and 'wall' was similar to the gaze data, possibly hinting at differing navigation strategies or perceptual preferences among participants.

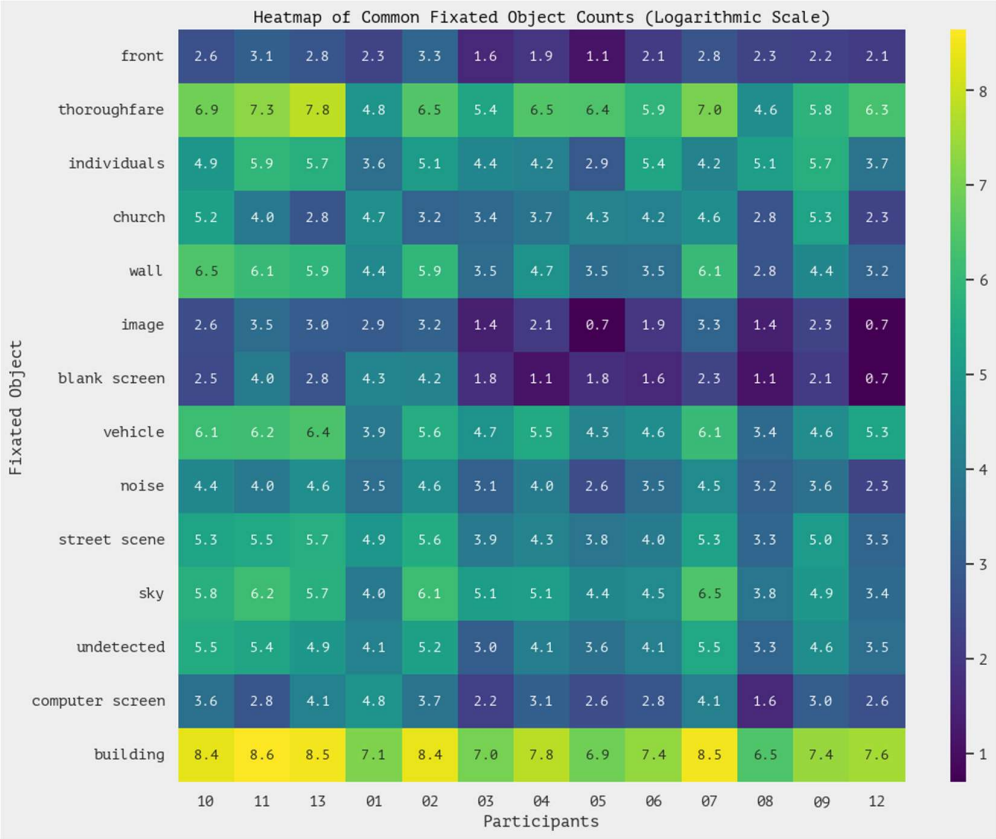


Figure 23 Heatmap of Common Fixated Objects Counts

High Frequency, High Fixation: Objects such as 'building', 'thoroughfare', and 'wall' are high both in occurrence and fixation. This might suggest that these objects are highly significant or salient in the environment and thus draw more visual attention.

High Frequency, Low Fixation: If there are objects with high frequency in the environment but low fixation, this could suggest that these objects are abundant in the scene but are not the focus of attention.

Low Frequency, High Fixation: Conversely, objects with low frequency but high fixation could be unique or special items in the environment that attract a disproportionate amount of attention.

Low Frequency, Low Fixation: Objects such as 'image' and 'front' that are low in both frequency and fixation might be of less significance to the subject or are less salient in the environment.

Through a comparison of these two heatmaps, an understanding can be gained concerning the relationship between an object's prevalence in the environment and its probability of attracting fixation. This can provide insights into the types of objects that are likely to capture attention in these settings and how the design of the environment may influence the focus of individuals' gaze.

4.1.2.2. Performance Correlation of Fixated Objects

These data present the correlations between task performance and the frequency of fixations on each object type. A positive correlation suggests an increase in fixation frequency is associated with a longer task completion time, reflecting lower spatial navigation performance as can be seen in the Table 5.

Building (0.987), vehicle (0.954): There is a very strong positive correlation between performance and fixation frequency on buildings. This means that participants who frequently fixated on buildings tended to take more time to complete the task, suggesting lower spatial navigation performance.

Object	*Correlation*
building	0.987
vehicle	0.954
street scene	0.902
wall	0.879
undetected	0.868
thoroughfare	0.861
sky	0.850
noise	0.829
image	0.827
front	0.808
individuals	0.544
blank screen	0.299
computer screen	0.235
church	0.043

Table 5 Common identified fixation correlation table

Street scene (0.902), wall (0.879), undetected (0.868), thoroughfare (0.861), sky (0.850), noise (0.829), image (0.827), front (0.808), individuals (0.544): For each of these objects, the positive correlations suggest that participants who fixated more frequently on these objects tended to take longer to complete the task. The strengths of these correlations vary, with most being moderate to strong. This suggests that there could be a meaningful relationship between the frequency of fixation on these objects and task performance.

It's crucial to understand that these correlations hint at relationships, but do not establish causality. Also, it should be noted that in this context, longer task completion times are associated with lower performance. Hence, positive correlations suggest a lower performance associated with increased fixation frequency. A negative correlation, if present, would suggest higher performance associated with increased fixation frequency.

4.2. Detailed Fixation Analysis

4.2.1. Fixation Metrics

In the process of analyzing eye-tracking data, three distinct parameters were calculated to offer insights into the focal points of the participants. These metrics, commonly used in eye-tracking studies, can provide answers to the primary research question.

Overall fixation duration quantifies the total amount of time a participant spent fixating on different Areas of Interest (AOIs) during the navigation task. It is expressed in seconds and highlights the number of seconds a participant focuses on, whether within or beyond the working memory capacity. The sensitivity of this measure extends to both actual memory load and processing load (Meghanathan et al., 2015). *Fixation count*, records the frequency with which a participant's gaze lands on a specific AOI to read text or examine an object. A higher fixation count can indicate an elevated level of interest or engagement with that area or object. *Average fixation duration* takes into account both the duration and count of fixations. A relatively higher average fixation duration for an object may suggest increased interest or cognitive engagement. This metric can be especially useful when analyzing objects or areas intended to draw the user's attention or deeper processing.

Given that the first two metrics are time-dependent, the average fixation duration was calculated to ensure a holistic understanding of the results. By incorporating both the duration and count of fixations, this metric offers a more detailed perspective on how participants interacted with various elements during the navigation task.

Overall Fixation Duration

As it can be seen in Figure 24, Participant 11 was noted to have the highest overall fixation duration at 720 seconds, closely followed by Participant 13 with 710 seconds. The extended time spent in fixation could indicate a higher level of information processing or possible navigational challenges. Conversely, it could also reflect heightened interest or engagement in the task.

Participant 8 recorded the briefest fixation duration, clocking in at 85 seconds. This could imply that this participant was particularly adept at processing visual information or perhaps not as thoroughly engaged with the task as others. However, given their notably swift and successful completion of the task, it seems more likely that Participant 8 demonstrated high efficiency in navigation.

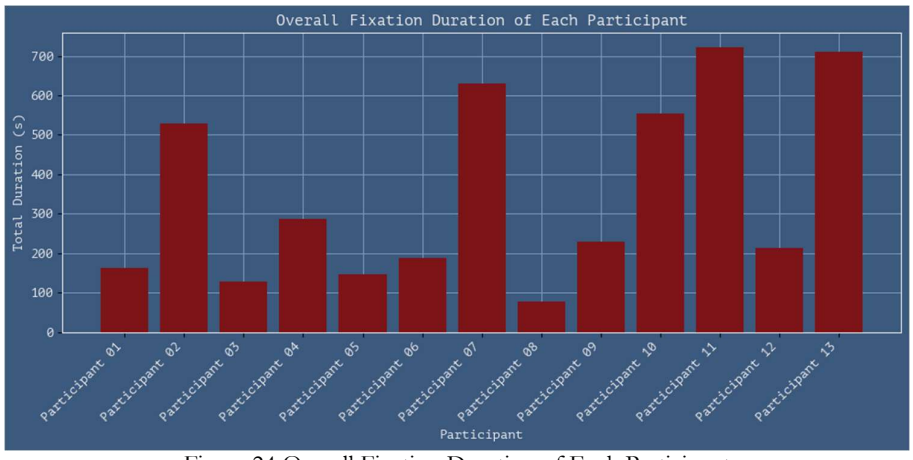


Figure 24 Overall Fixation Duration of Each Participant

Fixation Count

In Figure 25, it is shown that Participant 11 registered the highest number of fixations at 10,400, closely followed by Participant 13 with 9,950. This might indicate that these participants spent more time processing information or had more difficulty navigating the task, as evidenced by the increased number of fixations.

Conversely, Participant 8, despite having the briefest overall fixation duration, registered a relatively higher count of fixations (1,200) than some peers. This might suggest that their gaze moved more frequently, potentially indicating a different strategy of scanning or exploration.

On the other hand, Participant 1, with a moderate overall fixation duration of 180 seconds, recorded the fewest number of fixations at 2,000. This implies that this participant, while not spending an excessive amount of time fixating, did fixate for longer durations when they did. This could signify more time invested in processing each piece of visual information.

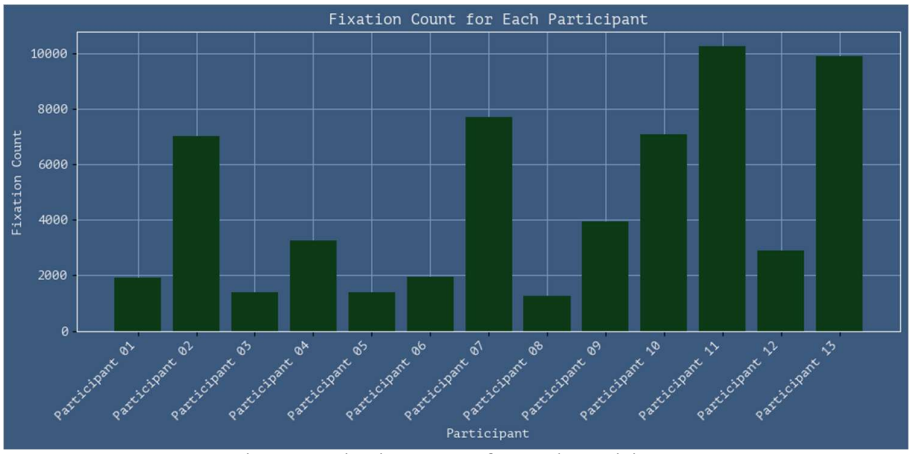


Figure 25 Fixation Count for Each Participant

Average Fixation Duration

As illustrated in Figure 26, the average fixation duration can serve as a metric of how participants process visual stimuli. Brief fixation durations may indicate swift visual processing, suggesting a strategy of rapid scanning. Conversely, extended durations could denote a more intensive engagement, indicative of deeper processing. With the longest average fixation duration of 107 milliseconds, Participant 5 appeared to engage more deeply with each visual element, hinting at a thorough, detail-focused approach. In stark contrast, Participant 9 exhibited the briefest average fixation duration of 58 milliseconds, suggesting a more superficial engagement with the visual stimuli. This might point to a strategy of rapid scanning, potentially beneficial for tasks requiring broader understanding.

Interestingly, Participant 7, despite exhibiting one of the longest total fixation durations and the highest fixation counts, recorded a relatively short average fixation duration of 82 milliseconds. This pattern could point towards a quick exploration strategy, with the participant frequently shifting focus across the visual environment. Similarly, Participant 8, who had the shortest total fixation duration, also recorded a below-

average fixation duration of 62 milliseconds. This pattern could be interpreted as efficient visual processing, with less overall time spent fixating and each fixation being relatively brief.

In contrast, Participant 1, with a moderate total fixation duration and the lowest fixation count, had a relatively lengthy average fixation duration of 85 milliseconds. This could suggest more in-depth engagement with each visual element. These observations suggest a diversity of visual processing strategies among participants. Some appear to prefer rapid scanning, while others seem to engage more deeply with the visual stimuli. Influences shaping these differences could range from individual cognitive styles to task specifics and the complexity of the visual environment.

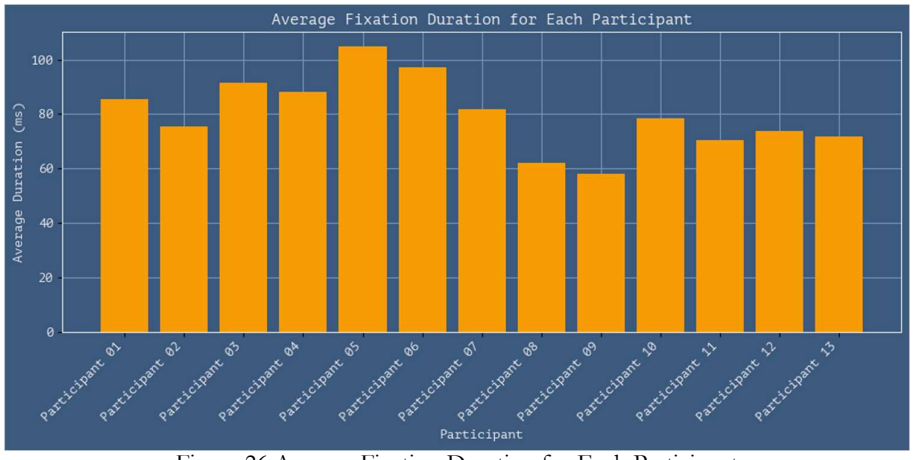


Figure 26 Average Fixation Duration for Each Participant

While the abovementioned metrics offer a generalized understanding of participants' fixation behavior and processing of their surroundings, additional analyses were performed to delve deeper into the fixation data. For the purpose of this study, participants were divided into two groups based on their task performance – 'Passed' and 'Failed'. The hypothesis tested was Hypothesis 2, "there is a significant difference in the metrics of fixation count, fixation duration, and average fixation duration between participants who successfully completed the task and those who did not."

4.2.1.1. Hypothesis Testing

To ensure that the data met the assumptions necessary for a t-test or Mann-Whitney U test, the Shapiro-Wilk test was first conducted to statistically evaluate the normality of the data. In this test, the null hypothesis posits that "the sample originates from a normally distributed population." If the p-value derived from this test is low, the null hypothesis can be rejected, suggesting that the data sample does not conform to a normal distribution.

Shapiro-Wilk Test	*Passed*	*Failed*
Avg. fixation duration (ms)	p-value: 0.9455	p-value: 0.3027
Overall fixation duration (s)	p-value: 0.0234	p-value: 0.1269
Fixation count	p-value: 0.0185	p-value: 0.2011

Table 6 Shapiro-Wilk Test Results

Average fixation duration (ms):
For both the Passed and Failed groups, the data is likely normally distributed (p-value > 0.05).
Overall fixation duration (s):
For the Passed group, the data does not appear to be normally distributed (p-value < 0.05), while for the Failed group, the data is likely normally distributed (p-value > 0.05).
Fixation count:
Similar to Overall Fixation Duration, for the Passed group, the data is not normally distributed (p-value < 0.05), but for the Failed group, it appears to be normally distributed (p-value > 0.05).

Given these results, there is a combination of normally and non-normally distributed data across both groups.

The Shapiro-Wilk test results reveal that the *Average fixation duration* demonstrates normality, while the two other metrics, *Fixation count* and *Overall fixation duration*, do not align with a normal distribution. Importantly, although the *Average fixation duration* conforms to normality, the sample size at hand is not large enough to adequately carry out a t-test. In the domain of eye-tracking research, smaller sample sizes are commonplace, necessitating the use of alternative tests that do not presume normality. As such, the Mann-Whitney U test, a non-parametric test, is often applied. This test enables a comparison between two separate groups, and its application has been validated across a variety of eye-tracking studies. (Cheng et al., 2022)

In the realm of this research, the performance on the navigational task between two distinct groups, namely, the ones who "Passed" and those who "Failed", is compared. The identical p-values for *Overall fixation duration* and *Fixation count* in the Mann-Whitney U test may be attributed to the limited sample size, especially in the 'Failed' group with only 3 observations. With such a small sample, the ranks in the data may align similarly for both measures, resulting in the same U statistic and p-value. This outcome reflects the underlying structure of the data in the context of a constrained sample size. The performance is assessed through three distinct metrics:

Mann-Whitney U Test	*p-value*	*U statistic*
Overall Fixation Duration (s)	0.0140	1.0000
Fixation Count	0.0140	1.0000
Avg. Fixation Duration (ms)	0.2867	22.0000

Table 7 Mann-Whitney U Test Results for Fixation Metrics

Overall fixation duration

The results present a U statistic of 1.0 and a p-value of 0.0140. The low U statistic indicates a substantial divergence between the groups, as in nearly all cases, a value from the "Passed" group ranks before a value from the "Failed" group when all data points are organized in order. The p-value (0.0140), being less than the conventional threshold for statistical significance (0.05), allows the rejection of the null hypothesis positing no difference between the two groups. Hence, it can be inferred that the "Passed" and "Failed" groups demonstrate significantly different *Overall fixation durations*.

Fixation count

In this case, the U statistic equals 1.0 and the p-value stands at 0.0140, mirroring the figures observed for the *Overall fixation duration*. This suggests that a significant difference exists between the two groups with respect to their *Fixation count*.

Average Fixation Duration

With respect to the *Average fixation duration*, a U statistic of 22.0 and a p-value of 0.2867 are observed. This higher U statistic indicates a less distinct difference between the groups that "Passed" and "Failed", as compared to the other metrics. Furthermore, the p-value exceeds 0.05, which leads to a failure to reject the null hypothesis. Consequently, no statistically significant difference can be discerned in the *Average fixation duration* between groups that "Passed" and "Failed" the navigational task.

From the given results, it can be inferred that the groups that "Passed" and "Failed" the navigational task significantly differ in terms of their *Overall fixation duration* and *Fixation count*, but not in their *Average fixation duration*.

Given the assumption that higher *Fixation counts*, and longer *Overall fixation durations* may signify greater difficulty with the task, these results suggest that participants who failed the task had a tendency to fixate more and for longer overall durations than those who passed. However, no significant difference in the average duration of each individual fixation was found between the groups that passed and failed.

These insights might suggest that efficiency in scanning, characterized by fewer fixations and shorter I, plays a crucial role in successfully completing the navigational task. This implies that participants who managed to quickly scan and process visual information, thereby needing fewer overall fixations and having shorter cumulative *Fixation durations,* were more successful in the task. This conclusion, if valid, has several implications. Firstly, it underscores the importance of the ability to rapidly and efficiently process visual information in navigational tasks. This could be particularly relevant in contexts where quick decision-making based on visual cues is vital.

Secondly, it highlights the fact that the length of individual fixations is not as critical in determining success. Despite the common belief that longer fixation durations may allow for deeper processing of visual information (Schwedes & Wentura, 2016), in this particular navigational task, it did not translate into better performance. It suggests that participants who fixated longer might have been overprocessing the visual information, leading to slower reactions and decisions, which could be detrimental in fast-paced, dynamic navigational tasks.

It should be noted that these interpretations are based on the specific context and task in this study. Other tasks or contexts might show different relationships between fixation characteristics and performance. As such, it's crucial to corroborate these findings with additional research, ideally with a larger sample size and diverse tasks to increase the generalizability of the results.

4.2.2. Scanpaths

In the context of this research, scanpaths are a valuable tool for understanding participants' visual exploration and behavior during the navigation task. They represent the sequence and spatial pattern of visual fixations and saccades (rapid eye movements between fixation points) made by the participants while interacting with the environment. (Davies et al., 2016) In the current research, scanpaths were continuously recorded from the start to the end of the navigation task. This comprehensive tracking led to

a large volume of trajectories, making detailed analysis challenging due to the complexity and density of the paths. However, this rich data also allows for insightful observations regarding the distribution of the participants' attention. While the specific trajectory details might be difficult to discern in such a busy visualization, we can still discern broader patterns or trends in attention allocation across the navigation environment. These overall trends provide valuable insight into areas where participants tend to focus their visual attention during navigation tasks.

The following sections provide a selection of participant scanpaths, illustrating some of the observed attention allocation trends. However, it should be noted that these are merely samples, and the underlying scene keeps changing as the person moves through the Google photospheres. These scan paths are within the person's field of view on the monitor, reflecting their navigational choices and gaze behavior as they interact with the dynamic environment. A complete collection of the scanpaths for all participants can be found in Annex 4.

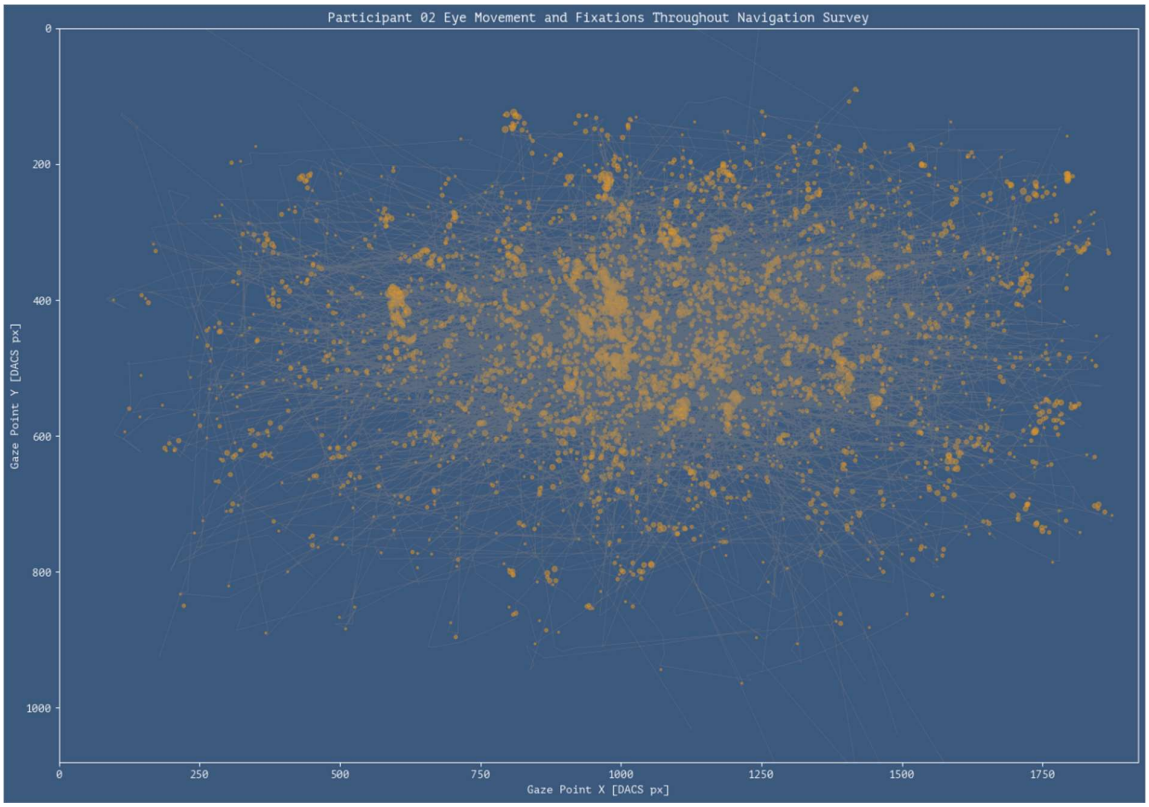


Figure 27 Scanpath of Participant 02

In the scanpath of Participant 02, Figure 27 fixations appear to be relatively homogeneously dispersed across the screen, indicating an evenly distributed pattern of attention allocation during the task. This broad attentional coverage might suggest an exploratory strategy, as the participant did not fixate intensely on any specific area. However, a moderate concentration of fixations can be observed towards the center of the screen. This central bias is a common finding in eye-tracking research, reflecting a general tendency to explore the central region of visual fields (Tatler, 2007). Nonetheless, in this case, the absence of any extreme focal points in the scanpath implies that Participant 02 did not consistently concentrate on a

single point or area in their field of view. This behavior might indicate a broad, but not in-depth, approach to gathering visual information during the navigation task.

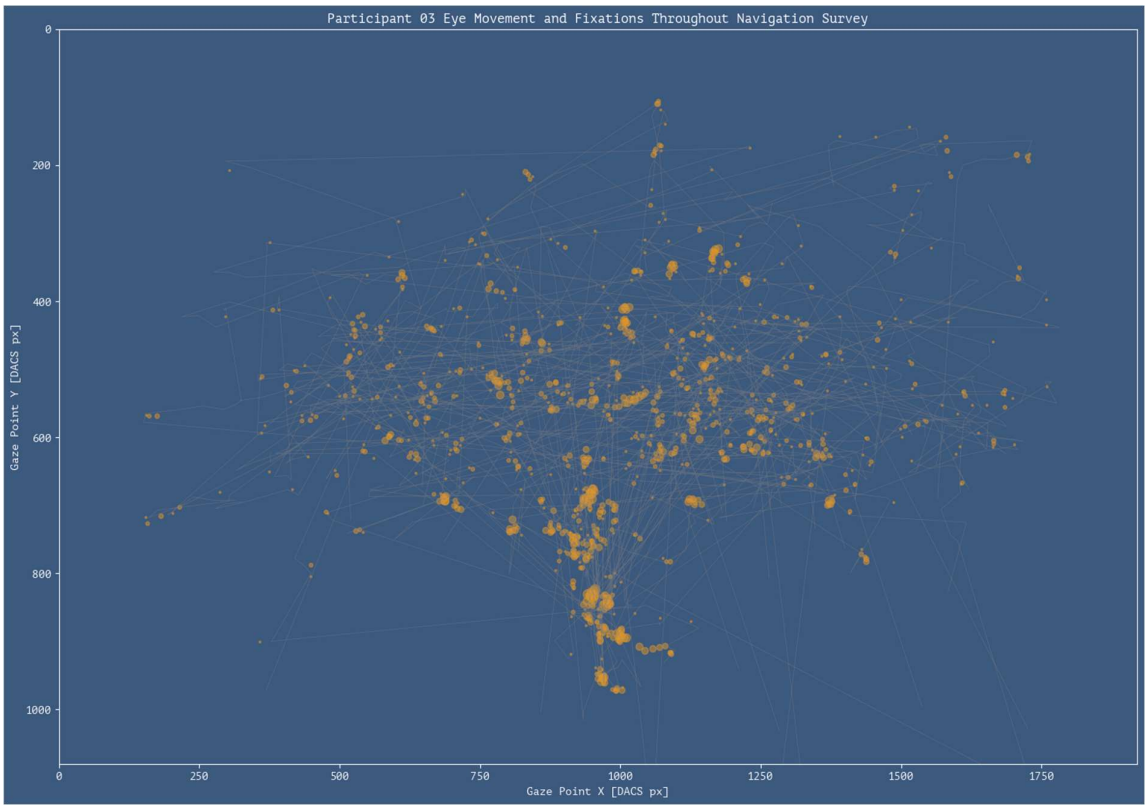


Figure 28 Scanpath of Participant 03

The scanpath of Participant 03, who performed exceptionally well by completing the task quickly, reveals a distinctive pattern. A substantial focus is evident on the center to lower part of the screen, where the pavement of the scene is predominantly situated. This heavy fixation on the thoroughfare suggests that the participant may have utilized this element of the urbanscape as a primary reference point or cue for navigation. Alternatively, they might have been looking for the user interface arrows that Google provides, using them as tools to guide their way through the task.

The trajectories, predominantly moving in and out from this central-lower region, further reinforce this interpretation. The participant's eye movements seem to "anchor" on the thoroughfare and then branch out to other parts of the scene, only to return to the thoroughfare again. This suggests an iterative visual strategy, perhaps continuously referencing the street while exploring other aspects of the environment.

The rest of the scanpath exhibits a more uniformly distributed pattern with no specific focal points. This part of the scanpath could represent a secondary level of information gathering, where the participant looks beyond the primary reference point (the thoroughfare) to gather additional visual cues from the wider urbanscape. It indicates that while the participant prioritizes certain features for navigation, they do not ignore the other aspects of the scene. This strategic approach might explain their success and efficiency in completing the navigation task.

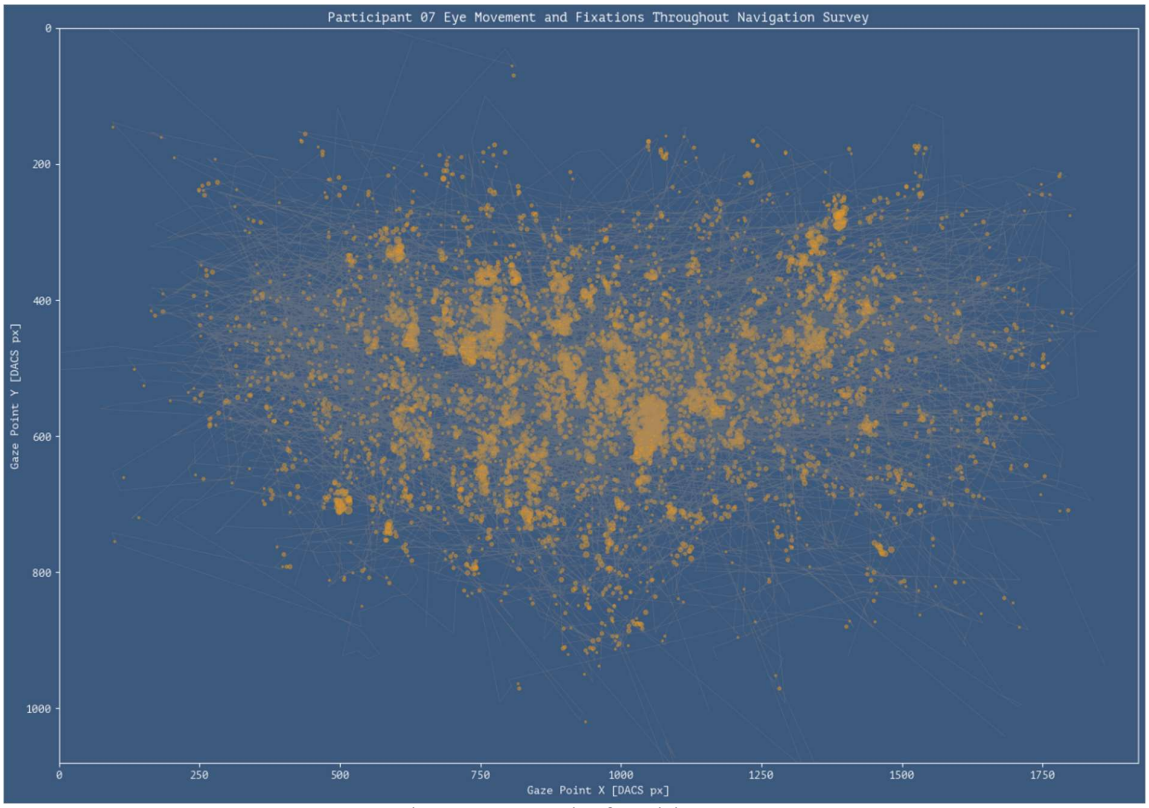


Figure 29 Scanpath of Participant 07

The scanpath of Participant 07, who completed the task but took a relatively long time, demonstrates a different pattern from those who performed well. The participant's fixations appear evenly distributed across the screen with no particular focal points. This dispersion may indicate a lack of a solid strategy or a clear reference point, contributing to the participant's less efficient performance.

Notably, a high number of trajectory lines are apparent within the scanpath. This abundance of lines signifies an elevated rate of saccades - rapid, jerky movements of the eyes as they shift focus between points. Such a high frequency of saccadic activity may be an indicator of the participant's restlessness or anxiety while navigating. In line with the think-aloud data, Participant 07 mentioned feeling lost multiple times during the task. The anxious, frantic searching for visual cues, as represented by the numerous saccades, may indeed be a response to this feeling of disorientation. This interpretation offers a plausible link between the subjective experience of feeling lost and the objective measures derived from the scanpath, reinforcing the complexity and emotional context of navigation tasks.

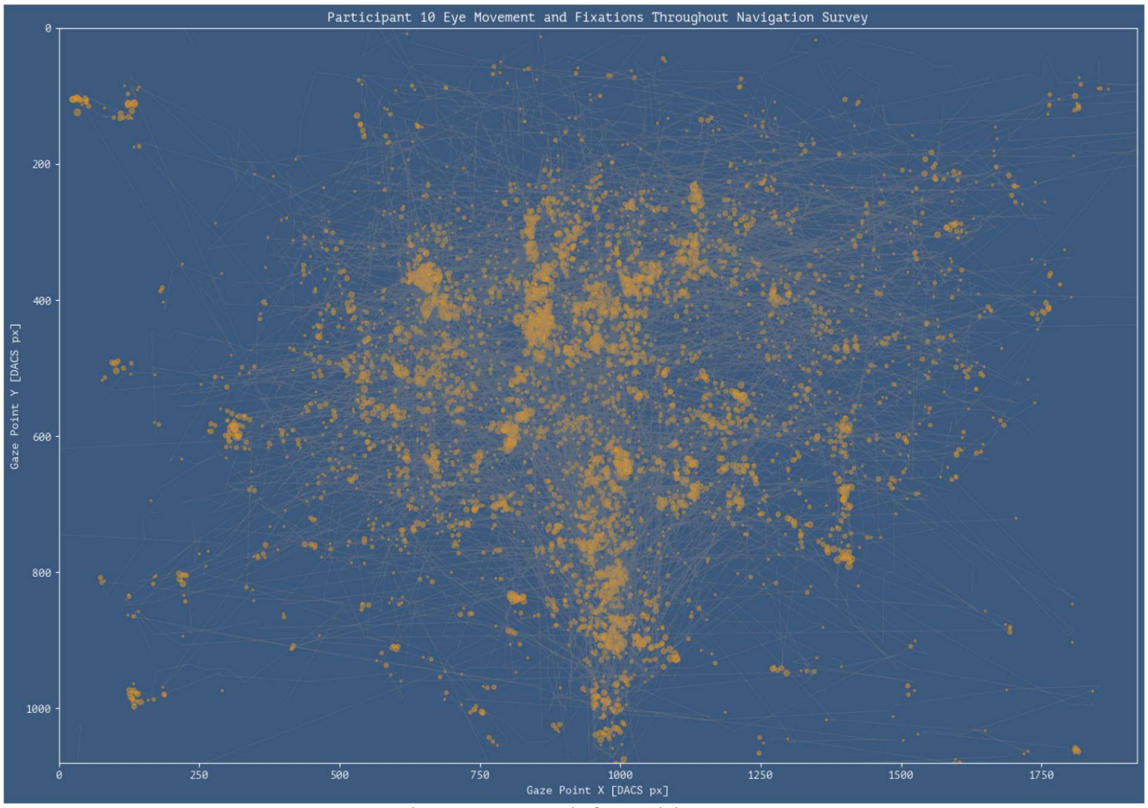


Figure 30 Scanpath for Paricipant 10

Participant 10's scanpath presents a slightly different pattern compared to others. This participant's fixations are widely scattered across the entire screen. While most participants tend to concentrate their attention within the central area, Participant 10 displays a high degree of fixation activity across all areas, including the corners of the screen. This behavior, combined with the participant's unsuccessful completion of the task, suggests a lack of a focused navigation strategy or difficulties in effectively utilizing visual cues from the environment.

A noteworthy aspect of this participant's scanpath is the apparent linear formation of fixations running from the lower-center towards the middle of the screen, which corresponds with thoroughfare in the scene. This may indicate a reliance on the pavement as a navigational reference, yet it does not seem to be sufficient to aid the participant in successful task completion.

Additionally, the scanpath exhibits trajectories that extend in all possible directions. Contrary to other participants who predominantly have trajectories moving outwards and then back towards the screen center, Participant 10's eye movements are significantly more disordered and unpredictable. This chaotic pattern may reflect a state of confusion or uncertainty, corroborating with the participant's unsuccessful task completion. This observation underlines the potential impact of the lack of a consistent and effective visual strategy on task performance.

4.2.3. Fixation Clustering

In this study, the use of Gaussian Mixture Model (GMM) for fixation clustering was extended to distinguish between successful and failed attempts at the navigation task. This allowed for a separate examination and comparison of the scanpaths between these two groups. The attention patterns, in the form of fixation clusters, were identified individually for participants who successfully completed the task and those who failed to do so. This distinction provided deeper insights into the variances in visual attention and navigation strategies between the successful and unsuccessful participants.

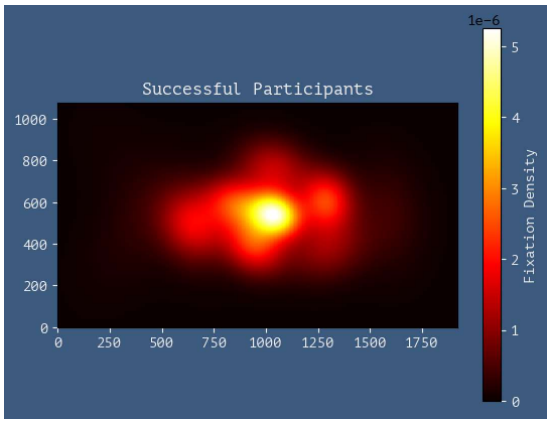


Figure 31 GGM for Successful Participants

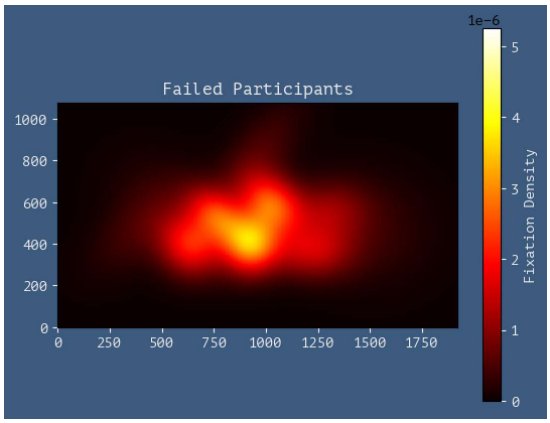


Figure 32 GGM for Failed Participants

The Intersection over Union (IoU) is a metric used to quantify the percentage overlap between two clusters. A higher IoU score indicates a larger overlap. In this case, the IoU of 0.683 suggests a substantial overlap between the fixation clusters of successful and unsuccessful participants, indicating that both groups exhibited similar patterns of visual attention during the navigation task. This result could also be reflective of the central bias in eye tracking, where participants have a natural tendency to focus on the center of the screen, regardless of their success in the task.

Metric	*value*
Intersection over Union:	0.6831610044313147
Centroid Distance:	8.43967318551894 px

Table 8 IoU and Centroid Distanc Results for GMM

Centroid distance, on the other hand, measures the distance between the centers (or 'centroids') of two clusters. In this analysis, the centroid distance is 8.44 px, implying that there is some divergence between where successful and unsuccessful participants fixated most frequently. However, given the high IoU, this divergence might not indicate a significant difference in overall attention distribution. In other words, both successful and unsuccessful participants tended to focus on similar areas of the urbanscape, but the most concentrated points of fixation (the centroids) were somewhat distanced.

To summarize, these results suggest that while there are some minor differences in exact fixation points, successful and unsuccessful participants exhibited largely similar patterns of visual attention during the navigation task. This finding provides valuable insights into the shared strategies used by participants, irrespective of their task success. Further investigation is needed to understand the subtleties that might differentiate successful navigation from unsuccessful attempts.

4.3. Hausdorff Distance Results

In the present analysis, deviation from the optimal path is quantified using the Hausdorff distance. This measure indicates the extent to which a participant deviated from the optimal path, defined here as the shortest route between start and end points of the navigation task. A lower Hausdorff distance signifies a closer adherence to the optimal path, whereas a higher Hausdorff distance reveals a greater deviation from that path. It's important to note, however, that a greater deviation from the shortest distance does not necessarily imply inefficiency. Participants could reach the destination very efficiently through an alternate path, such as by using a parallel street. Thus, the use of Hausdorff distance, while valuable, may not fully encapsulate the complexity of individual navigation strategies in the task.

The analysis of Hausdorff distances reveals varying degrees of path efficiency among the participants. As can be seen in the Figure 33 and Figure 34 Participants 03 and 09 demonstrate the smallest Hausdorff distances (71.49m and 71.55m respectively), indicating that they adhered more closely to the optimal path compared to other participants.

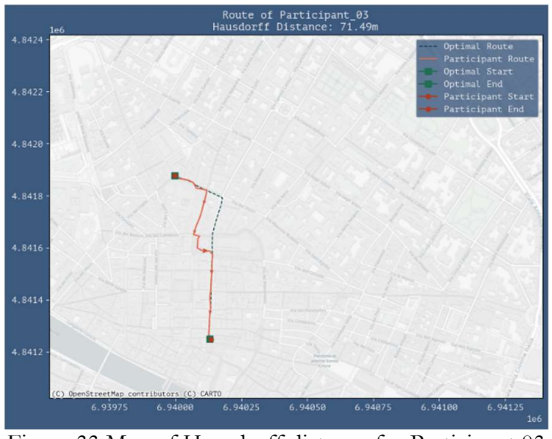


Figure 33 Map of Hausdorff distance for Participant 03

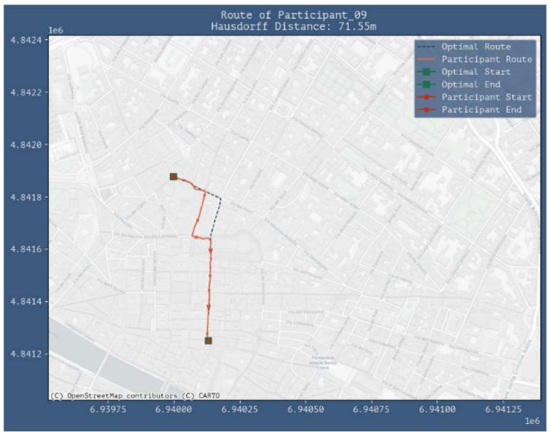


Figure 34 Map of Hausdorff distance for Participant 09

Participants 06 and 08 also display relatively small Hausdorff distances, which suggests they too followed efficient paths that deviated minimally from the optimal trajectory.

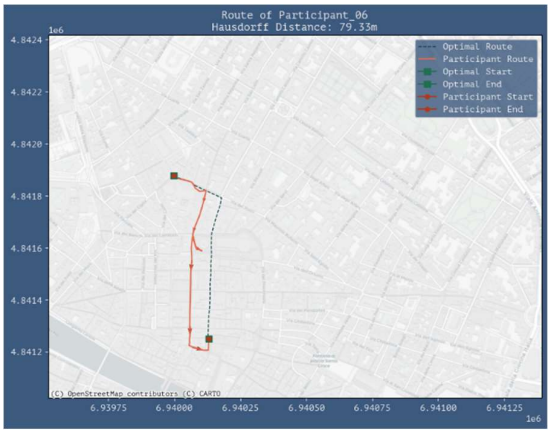


Figure 35 Hausdorff Distance for Participant 06

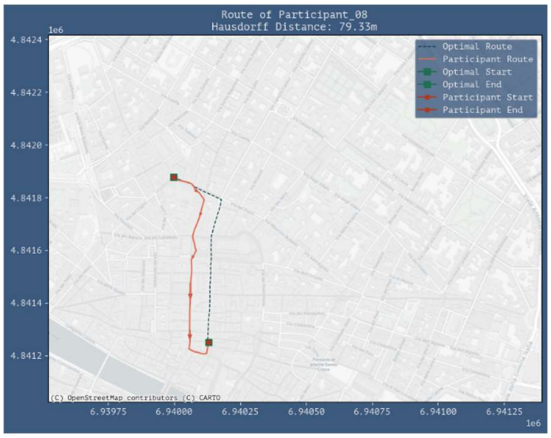


Figure 36 Hausdorff Distance for Participant 08

For Participants 01, 04, and 05, their Hausdorff distances were moderate. For Participant 04, the Hausdorff distance was moderate, despite taking a lot of back-and-forth movements in the pathfinding. This indicates that, although there were many directional changes, the participant did not deviate significantly from the optimal route.

Figure 37 Hausdorff Distance for Participant 01

Figure 38 Hausdorff Distance for Participant 04

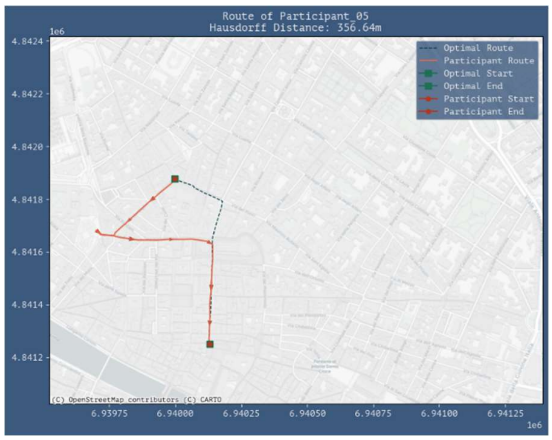


Figure 39 Hausdorff Distance for Participant 05

Figure 40 Hausdorff Distance for Participant 12

In contrast, Participants 02, 07, 11, and 13 recorded high Hausdorff distances, suggesting they significantly deviated from the optimal path and their chosen routes were likely less efficient.

Figure 41 Hausdorff Distance for Participant 02

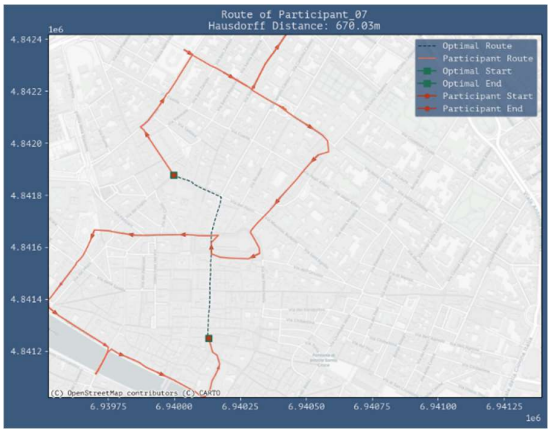


Figure 42 Hausdorff Distance for Participant 07

Figure 43 Hausdorff Distance for Participant 11

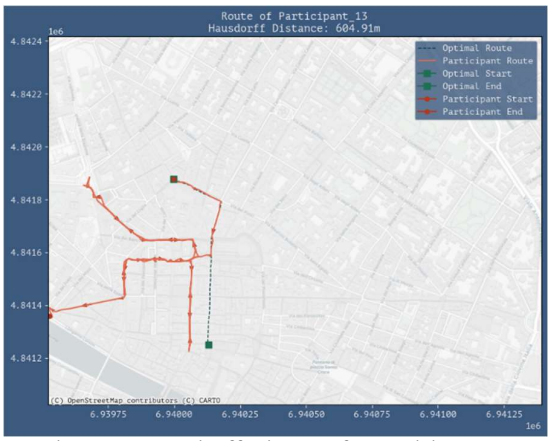


Figure 44 Hausdorff Distance for Participant 13

Most notably, Participant 10 recorded a Hausdorff distance of 1240.18, the largest among all participants. This high value indicates a substantial deviation from the optimal path, hinting at a possibly inefficient navigation strategy employed by this participant.

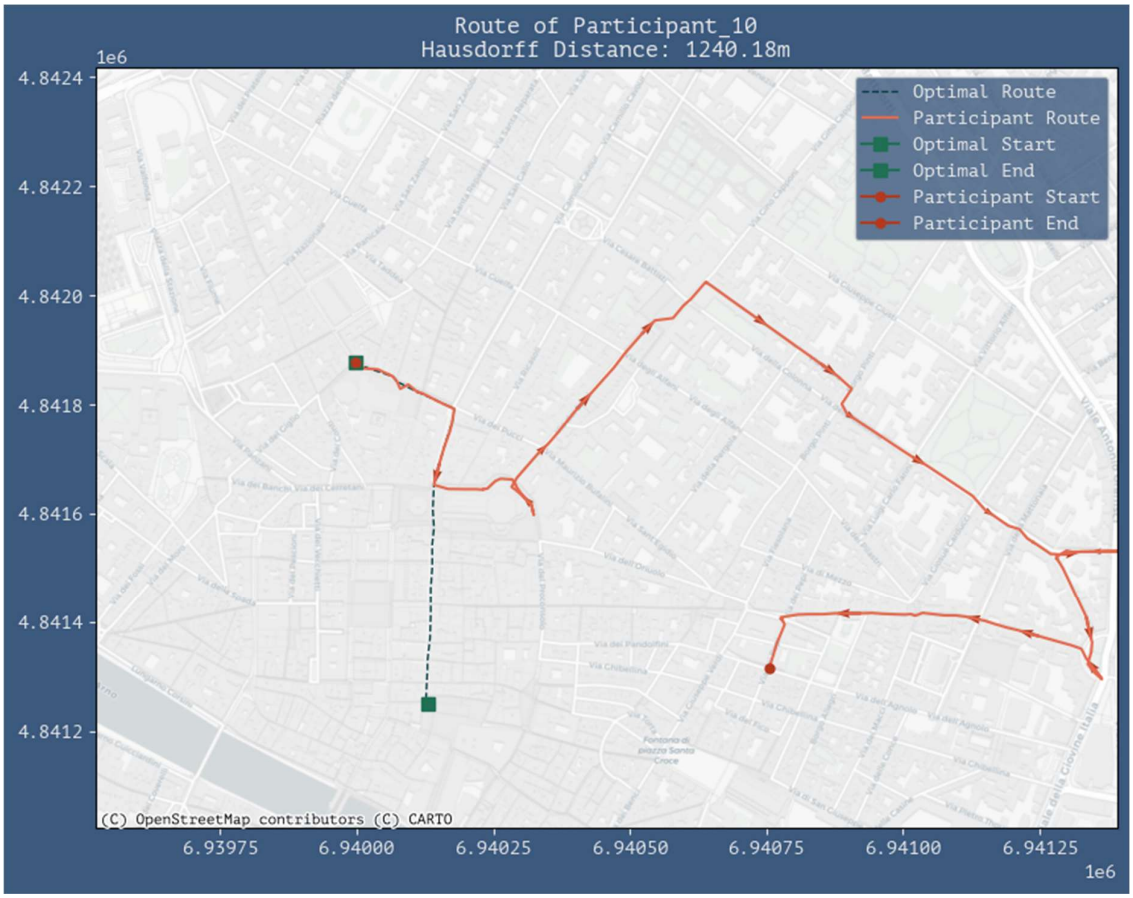


Figure 45 Hausdorff Distance for Participant 10

In the Table 9, the directed Hausdorff distances are calculated in two directions: the forward distance from the optimal route to the participant's route, and the backward distance from the participant's route to the optimal route. In this study, the mean Hausdorff distance is found to be equal to the mean forward Hausdorff distance, at 421.19m. This occurs because the final Hausdorff distance for each comparison is taken as the maximum of the forward and backward distances, and in this case, the forward distance is greater for each comparison. The term 'forward' refers to the directed distance from the optimal route to the participant's route, while 'backward' refers to the directed distance from the participant's route to the optimal route. These directional measures provide a comprehensive comparison, capturing the dissimilarity in both directions, which is particularly relevant in the comparison of routes. The statistical analysis reveals certain patterns in the participants' navigational abilities:

Metrics	*Value*
Mean Hausdorff distance	421.19m
Mean forward Hausdorff distance	421.19m
Mean backward Hausdorff distance	157.92m
Standard deviations of Hausdorff distances	338.82m
Standard deviations of forward Hausdorff distances	338.82m
Standard deviations of backward Hausdorff distances	97.45m

Table 9 Hausdorff Distance Descriptive Statistics

The average Hausdorff distance, which measures deviation from the optimal path, is 421.19m for all participants. This indicates that, on average, participants' paths deviated by 421.19m from the most efficient route. However, the standard deviation of Hausdorff distances is relatively high at 338.82m, suggesting considerable variation among participants' navigation skills. This large standard deviation implies a wide disparity in path selection, with some participants closely aligning with the optimal path and others veering significantly off course.

Interestingly, the mean forward Hausdorff distance, measuring from the optimal route to the participant's route, is equal to the mean Hausdorff distance at 421.19m. The mean backward Hausdorff distance, measuring from the participant's route to the optimal route, is significantly lower at 157.92m. This discrepancy suggests that the dissimilarity is more pronounced when considering the distance from the optimal route to the participant's route. The standard deviations for forward and backward Hausdorff distances stand at 338.82m and 97.45m, respectively, indicating a higher degree of variation in the participants' alignment with the optimal route in the forward direction. This could reflect differing levels of proficiency or strategy among the participants in navigating towards the optimal path.

4.3.1. Hypothesis testing

4.3.1.1. Mann-Whitney U Test for Hausdorff Distance

The Mann-Whitney U test has been applied to assess the relationship between Hausdorff distance and the performance outcome of the participants (either 'passed' or 'failed'). The U statistic of 2.0000 suggests a distinct difference between these two groups. A smaller U statistic indicates a more significant difference between the two groups. The p-value is 0.0344, which is less than 0.05, indicating that there is a statistically significant difference in Hausdorff Distance between participants who 'Passed' and those who 'Failed'. We can infer that the 'Passed' and 'Failed' groups follow significantly different paths from the optimal route.

The results obtained from the Mann-Whitney U test suggest a significant relationship between Hausdorff distances and task success. This affirms the hypothesis that adherence to the optimal path is associated with task performance. Simply put, lower Hausdorff distances, implying less deviation from the optimal route, are observed more frequently among participants who successfully completed the task. Conversely, higher Hausdorff distances, indicating greater deviations, are observed more often among those who did not complete the task successfully.

These findings underscore the significance of route optimization in navigational tasks. Such implications could extend to real-world applications, such as in the design of navigational tools, the establishment of training programs, and the development of strategies to improve navigational efficiency across various contexts, from driving to trekking to traversing intricate facilities.

Notably, while the statistical test does highlight a significant difference, it does not elucidate the magnitude or practical significance of the observed difference. The correlation between Hausdorff distance and task performance, while statistically significant, does not necessarily imply that all individuals with higher Hausdorff distances will fail, or that all with lower Hausdorff distances will succeed. The existence of other influential factors should be considered, and the possibility of individuals overcoming deviation through problem-solving strategies or other skills cannot be dismissed.

4.3.1.2. Spearman's Rank Correlation

The selection of Spearman's Rank Correlation as the statistical method for this dataset and research objectives was driven by its unique advantages suitable for the specific characteristics of the data at hand. As a non-parametric statistical procedure, Spearman's correlation is ideal for this study due to the relatively small sample size, suggesting that the data may not strictly adhere to a normal distribution.

This method's distinctive sensitivity to data rankings, not absolute values, proves beneficial when examining relationships between variables measured on different scales or units, such as the Hausdorff distance and Overall Fixation Duration in the present study. Furthermore, unlike Pearson's correlation, Spearman's correlation has the capacity to identify both linear and non-linear monotonic relationships, an ability that is particularly advantageous given the possible non-linear relationships between variables in the dataset.

Another essential factor in choosing Spearman's correlation is its robustness against outliers. Since the computation of Spearman's correlation relies on data ranks rather than raw data, it demonstrates stronger resistance against potential outlier values, thereby enhancing the reliability of the analysis. Thus, in light of the nature and objectives of the current study, Spearman's Rank Correlation was considered the most judicious selection, bearing in mind that the choice of statistical test largely depends on the dataset's specific characteristics and research goals.

Hausdorff Distance and Overall Fixation Duration

The correlation coefficient is 0.7318, which is a strong positive correlation. This suggests that as the Hausdorff Distance increases (meaning participants deviate further from the optimal path), the Overall Fixation Duration also increases. The p-value is 0.0045, which is less than 0.05, indicating a statistically significant correlation. This result implies that there is a significant relationship between the path deviation and the overall fixation duration.

Hausdorff Distance and Avg. Fixation Duration

The correlation coefficient is -0.0770, suggesting a weak negative correlation, indicating that as the Hausdorff Distance increases, the Average Fixation Duration slightly decreases. However, considering the p-value of 0.8025, which is larger than 0.05, we fail to reject the null hypothesis. This result means there's no statistically significant correlation between Hausdorff Distance and Average Fixation Duration. Hence, the extent of deviation from the optimal path doesn't significantly impact the average duration of fixations.

Hausdorff Distance and Fixation Count

The correlation coefficient between Hausdorff Distance and Fixation Count is 0.6713, suggesting a moderately strong positive correlation, indicating that as the Hausdorff Distance increases, the Fixation Count also increases. The p-value is 0.0120, which is less than 0.05, indicating a statistically significant correlation. While this correlation implies a significant association between the amount of deviation from the optimal path and the number of fixations, it is worth considering that the correlation may also be influenced by other factors. Specifically, as pointed out by the supervisors, this correlation might be expected due to the time taken to complete the task and the opportunities to look around as a result of

wandering more. A person who wandered more would naturally have had more time and street space to look around, which may contribute to the observed correlation. Therefore, while the correlation is statistically significant, the interpretation should be made with an understanding of these underlying dynamics that might also be at play.

Additionally, this deviation is significantly correlated with both the Overall Fixation Duration and the Fixation Count. This implies that participants deviating more extensively from the optimal path tend to demonstrate a higher number of fixations and a lengthier overall fixation duration. Such observations could be interpreted as an indication that the navigational task becomes increasingly demanding for participants straying from the optimal route, prompting more frequent and longer fixations as they attempt to navigate.

However, it is worth noting that there was no substantial difference observed in the Average Fixation Duration between participants deviating more from the optimal path and those staying closer. This suggests that while the total duration and number of fixations tend to increase with greater deviation, the duration of individual fixations remains relatively consistent. This may indicate that as participants face more complex navigational tasks, they tend to examine their environment more frequently and for longer overall durations, but without extending the duration of individual fixations. This trend might suggest a heightened need for environmental scanning or surveying, rather than an increased depth of processing or interpretation at each point of interest.

These findings collectively provide support for the Hypothesis 3, suggesting that the way participants visually engage with their environment (as indicated by fixation metrics) is indeed related to the efficiency of their navigational choices (as indicated by the deviation from the optimal path).

4.4. Analysis of Think-Aloud Responses

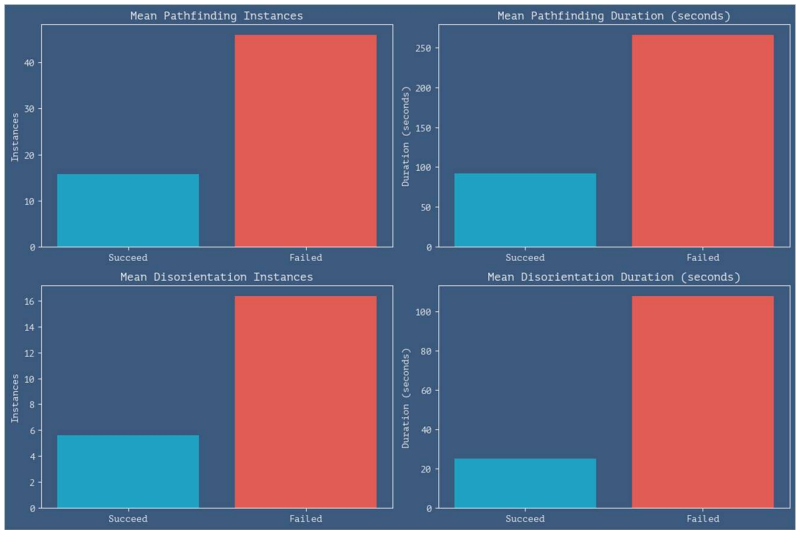
Analyzing the table provides valuable insights into the relationship between participants' performance (Succeed/Failed), their pathfinding activities, and instances of disorientation during the think-aloud tasks.

Participant	*Group*	*Pathfinding Instances*	*Pathfinding Duration (s)*	*Disorientation Instances*	*Disorientation Duration (s)*
01	Succeed	6	54	3	24
02	Succeed	22	141	16	77
03	Succeed	14	93	2	13
04	Succeed	17	78	11	39
05	Succeed	13	66	4	30
06	Succeed	9	78	1	4
07	Succeed	28	148	14	50
08	Succeed	9	65	0	0
09	Succeed	19	105	4	12
10	Failed	24	252	12	89
11	Failed	55	222	13	52
12	Succeed	20	92	1	4
13	Failed	59	325	24	182

Table 10 Think-Aloud Code Counts and Durations

Pathfinding Instances and Duration: It's observed that participants who failed the task generally had more pathfinding instances and spent longer periods on pathfinding compared to those who succeeded. Participant 11 and 13, who failed the task, had the highest number of pathfinding instances (55 and 59, respectively) and spent the longest time on pathfinding activities (222 and 325 seconds, respectively). This suggests a possible difficulty in understanding or navigating the task environment, which may have contributed to their failure.

Disorientation Instances and Duration: Similarly, participants who failed the task had a higher frequency of disorientation instances and spent a more extended time disoriented. This is particularly noticeable for Participant 13, who experienced 24 disorientation instances with a total duration of 182 seconds - the highest across all participants. This indicates a strong relationship between the frequency and duration of disorientation and task failure. In contrast, Participant 8 who successfully completed the task had no disorientation instances, suggesting a clear understanding and smooth navigation of the task environment.



Interplay between Pathfinding and Disorientation: Participants who struggled with pathfinding also seemed to experience more disorientation. The correlation between high pathfinding instances and high disorientation instances, particularly among those who failed the task, suggests these two factors may work in tandem, affecting a participant's ability to succeed in the task.

Successful Participants' Behavior: Successful participants generally had fewer pathfinding and disorientation instances, and they spent less time on these activities. However, it's worth noting that the relationship between these variables and success is not straightforward. For instance, Participant 7 succeeded despite having 28 pathfinding instances (the highest among those who succeeded) and spending 148 seconds on pathfinding. This could indicate other factors at play, such as the effectiveness of the pathfinding strategies used, rather than just the number of instances or duration.

Spearman's rank correlation	*value*
Pathfinding and Disorientation Instances	0.7959
Pathfinding and Disorientation duration	0.7342

Table 11 Spearman's rank correlation between think-aloud's codes

As shown in Table 11, the Spearman's rank correlation coefficients between instances of pathfinding and disorientation, and between pathfinding duration and disorientation duration, were found to be 0.7959 and 0.7342, respectively. These values reveal intriguing relationships among the measured variables. The strong positive correlation of 0.7959 between instances of pathfinding and disorientation indicates that participants who frequently engaged in pathfinding also experienced more instances of disorientation. This could reflect a complexity in navigation or perhaps a link between the very act of pathfinding and the occurrence of disorientation. Similarly, the moderately strong positive correlation of 0.7342 between pathfinding duration and disorientation duration suggests that participants who took longer to find their paths also tended to be disoriented for more extended periods. This relationship might imply that the longer time spent in pathfinding is associated with an increase in confusion or uncertainty during navigation. Both correlations highlight a complex interplay between pathfinding and disorientation, whether in instances or duration, revealing a nuanced understanding of participants' navigation skills and strategies.

This analysis underscores the importance of considering both the quantitative and qualitative aspects of task navigation. While the number of instances and the time spent on pathfinding and disorientation can provide a quantitative measure of participants' struggles, qualitative analysis might reveal more about the nature of these struggles, the strategies used to overcome them, and why these strategies were or were not effective.

Though the standalone analysis of "think-aloud" results may not highlight specific problems, further investigation is undertaken to provide a clearer understanding. By combining "think aloud" data with other information, the following sections aim to draw more comprehensive insights. The purpose of this integrated approach is to yield a richer interpretation of the results.

4.5. Comprehensive Data Fusion: Integrating Spatial Data, Eye-Tracking, and Think-Aloud Data

The analysis and interpretation conducted in this research primarily involved visual methods, comprising several datasets to deliver a detailed, comprehensive understanding. These datasets were mainly used to capture and analyse specific moments during a navigational task where participants paused to assess their environment. Each participant's route was visualized in a three-dimensional plot to enhance the distinction of observation points. This graphical representation utilized the X and Y axes for geographic coordinates and the Z axis for normalized fixation duration at each point. Normalization in this context ensures that fixation durations across various observation points are standardized, allowing for easy comparison and interpretation. By transforming the raw durations into relative values, the process emphasizes the patterns or trends in participants' behaviors rather than the absolute values, providing a clearer understanding of the underlying phenomena. Essentially, each bar in the plot represented a unique observation point.

Furthermore, each of these observation points was enriched with a stacked bar chart. This chart illustrated the duration of fixation on various objects at each point. While an attempt was made to include all objects from the navigation task in the plots, certain objects might not be clearly visible due to limitations in visualization.

A notable analytical challenge was presented when participants chose to retrace their steps along the same path. This resulted in overlapping lines and data points, thereby obscuring the clarity of individual paths

and associated think-aloud data. Despite this recognized limitation, solutions to this issue are beyond the scope of the current research.

Another layer of the analysis encompassed the think-aloud data, adding a further dimension to the understanding of the participants' experiences. For this part of the analysis, the verbal responses were classified into two categories - pathfinding and disorientation. Represented by the colours green and red respectively, these categories were included in all of the plots. However, it is important to note that the spatial accuracy of these codes may not be completely precise due to participants' delay in verbal response or their inability to articulate their exact thoughts.

Detailed plots for each participant are provided in Annex 5, with a select few discussed below for better understanding.

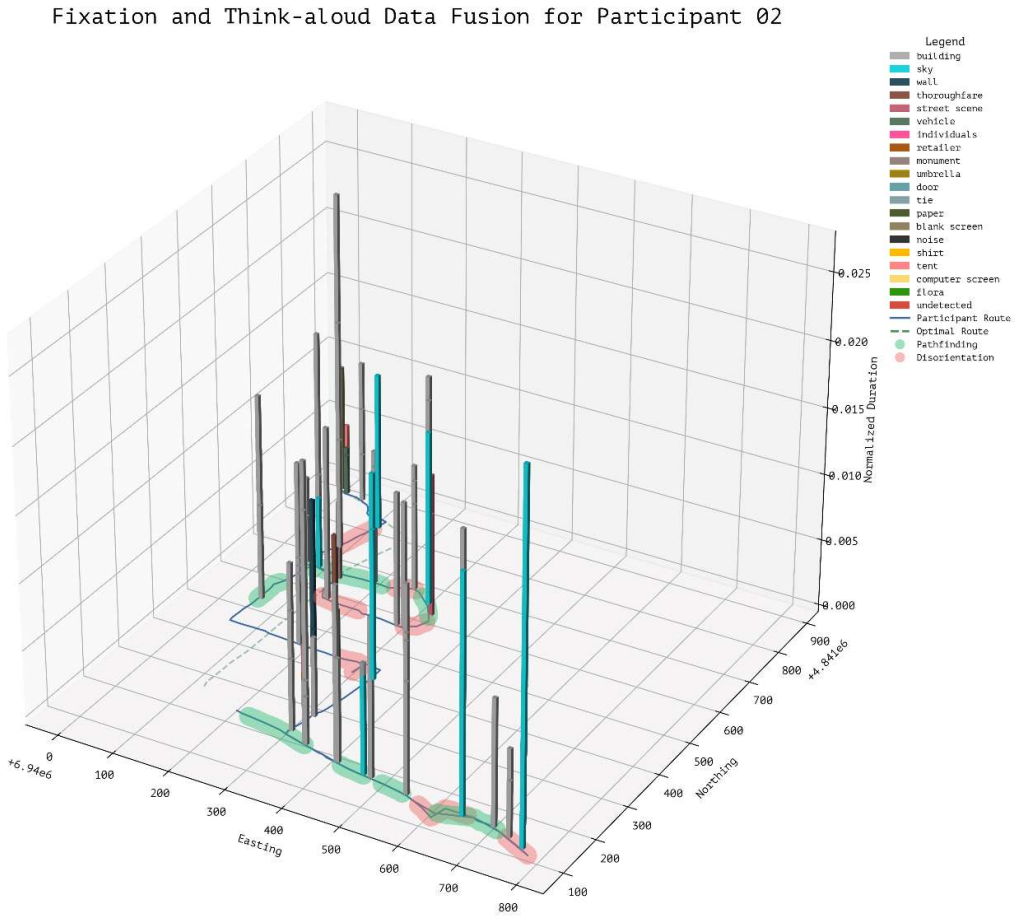


Figure 46 Fixation and Think-aloud Data Fusion for Participant 02

In Figure 46, Participant 02's route shows increased fixation duration at the start of the survey. This suggests that the participant was gathering more information about their environment before proceeding. There was a significant amount of pathfinding at the intersections and around the cathedral in Florence, which is noticeable as a loop in the middle of the map around the coordinates 600, 250. This participant exhibited a tendency to circle around the cathedral while trying to locate the third landmark. The

participant felt disoriented at this point, as reflected by the disorientation codes recorded. The participant also retraced their steps after entering an alley, marking another instance of disorientation.

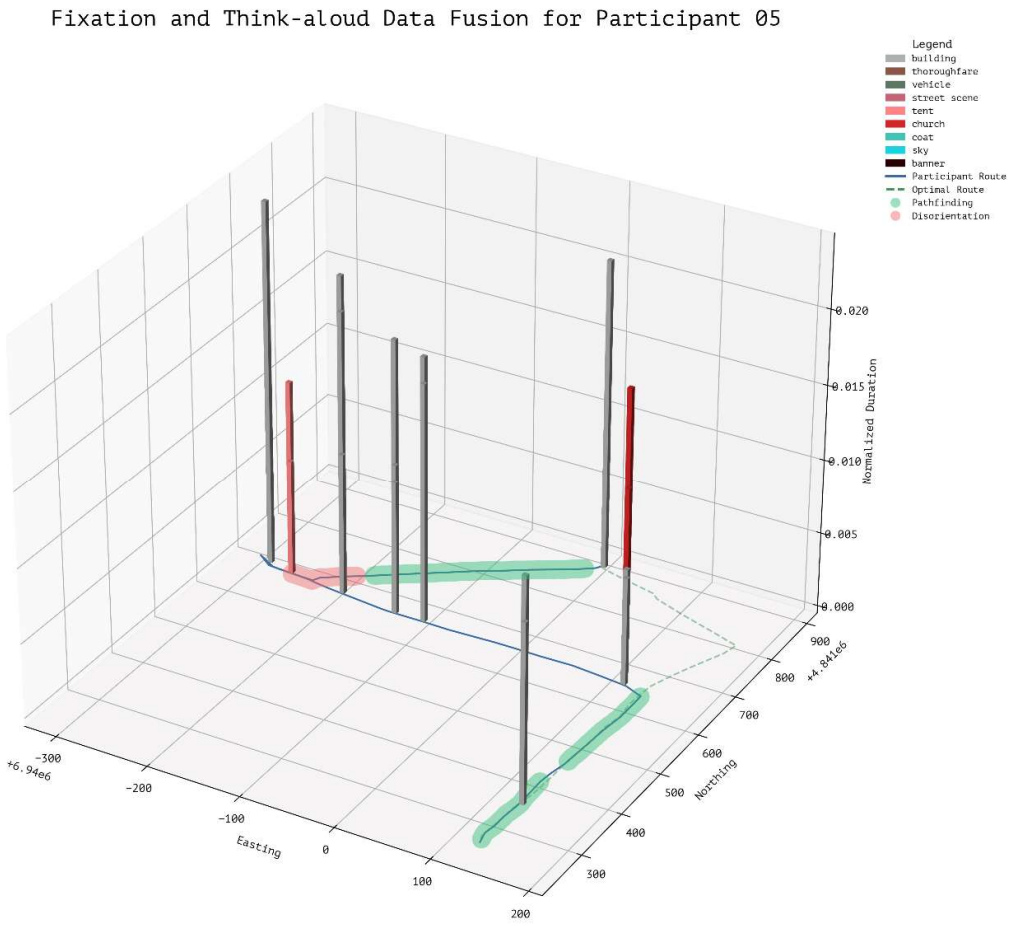


Figure 47 Fixation and Think-aloud Data Fusion for Participant 05

In Figure 47, it is evident that Participant 05 followed a relatively short path to the endpoint. Two distinct features are noticeable in this path: first, there is an observation point where the participant appears to have felt disoriented, fixating on the street scene for an extended period, likely in search of a visual cue. Second, close to the cathedral of Florence at a turning point in the route, the participant fixated for a more extended duration on the church object, specifically the cathedral of Florence. This behavior suggests an effort to recognize the landmark and navigate towards the final destination.

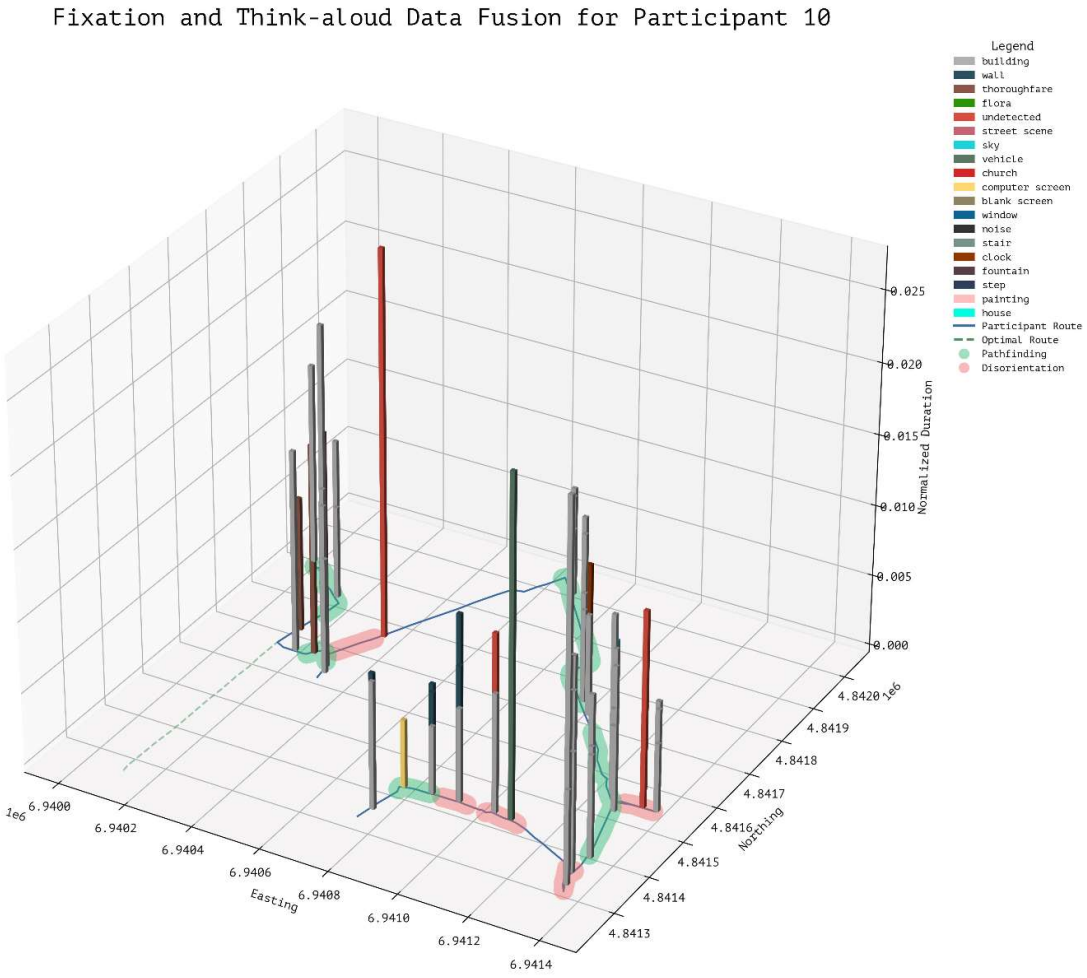


Figure 48 Fixation and Think-aloud Data Fusion for Participant 10

Participant 10's performance, depicted in Figure 48, showed a tendency to take straight routes until reaching boundaries, after which the participant relied more heavily on decision-making at intersections. This change in strategy, marked by a visible increase in disorientation codes, coincided with the participant feeling increasingly lost.

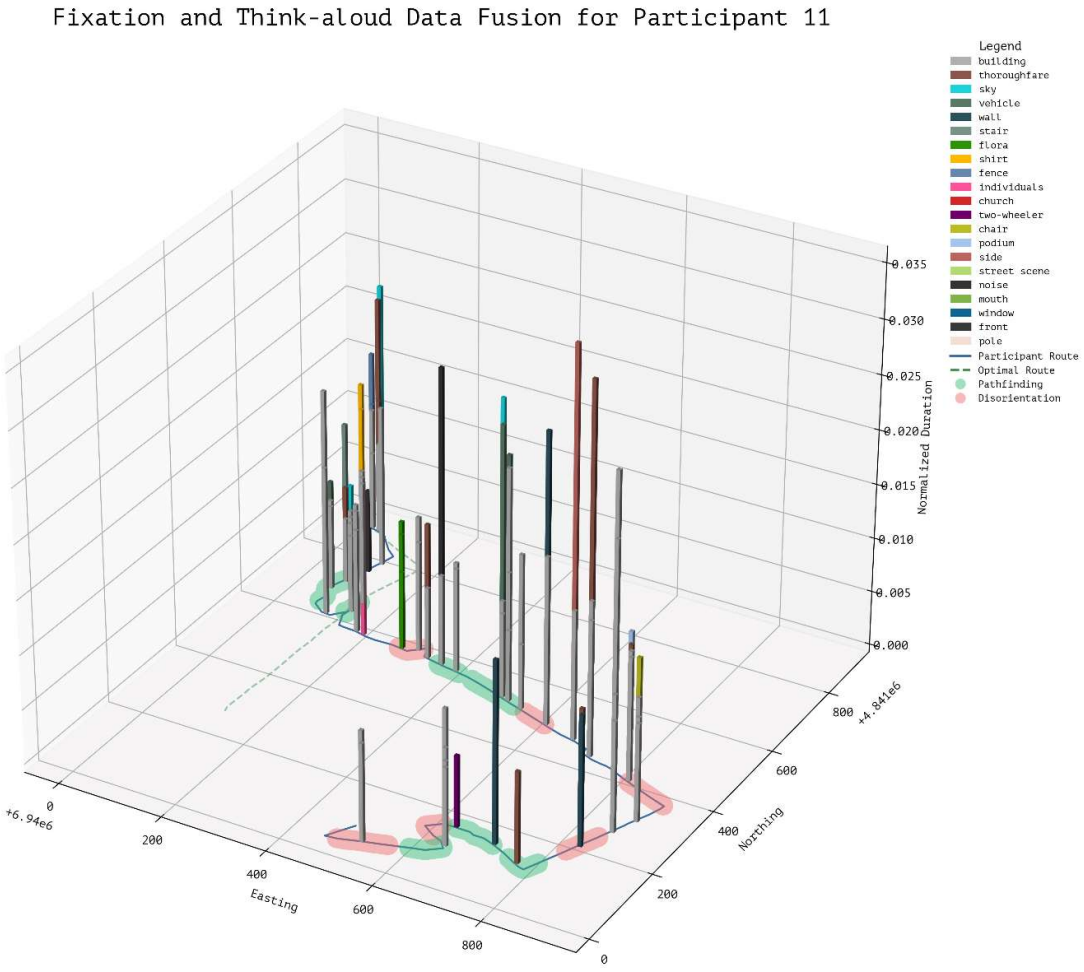


Figure 49 Fixation and Think-aloud Data Fusion for Participant 11

In Figure 49, Participant 11 took a different approach. The participant took the time to understand the environment, pausing frequently and demonstrating varied fixations on different objects at observation points. However, as the participant got closer to the end of the task, the participant felt increasingly disoriented and was unable to complete the task.

Participant 13, represented in Figure 50, exhibited a clear pattern of trying out different routes until the participant felt disoriented, at which point the participant would decide to either change the route or retrace the steps. As the task progressed, a noticeable decrease in fixation duration suggests a growing sense of fatigue or confusion.

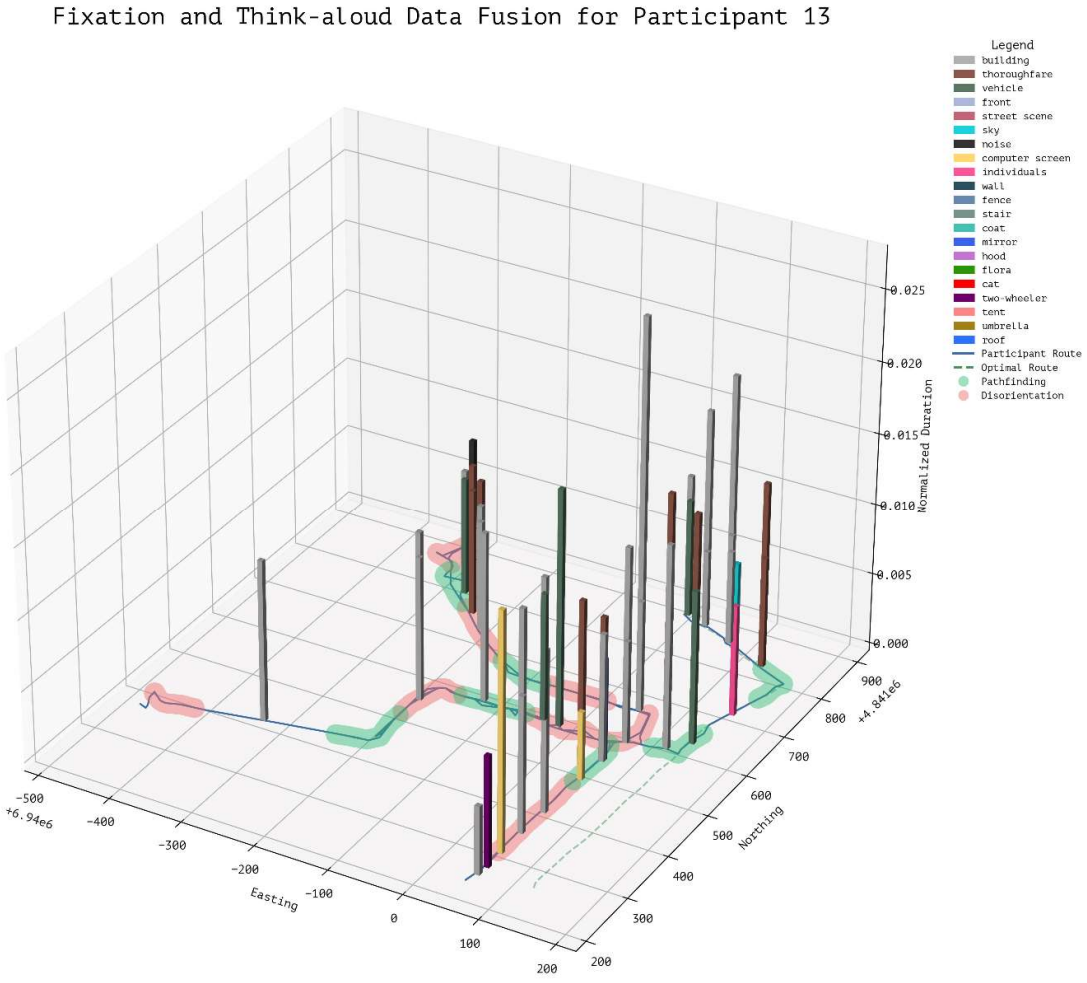


Figure 50 Fixation and Think-aloud Data Fusion for Participant 13

To conclude, this visual analysis provides valuable insights into how participants navigated their routes, their decision-making processes, and the challenges they encountered. These findings underscore the importance of examining spatial and temporal data to better understand human navigation behaviours, thereby contributing to ongoing research in this field.

5. CONCLUSION AND FUTURE WORK

5.1. Conclusions

n this research, an integrated approach was employed to gain a comprehensive understanding of visual behavior and navigation strategies in a navigation task setting. The core of this methodology involved the use of eye-tracking data, interpreted in conjunction with semantic segmentation, allowing for the identification and analysis of fixated objects in the urbanscape. In addition to visual analysis, the study incorporated spatial consideration. The following sections provide responses to the research questions posited at the beginning of this study.

1. What are the urbanscape objects that pedestrians tend to fixate on during the process of self-navigation in an unfamiliar urban environment with a nearby mapped destination, and without defined routes or navigational tools?

The primary focus of this research was to identify the objects that individuals fixate on while navigating an unfamiliar urban environment. Buildings emerged as the most commonly observed and fixated objects, suggesting their pivotal role as navigational guides in unfamiliar territories. However, individual variations in navigational strategies were uncovered, with different participants fixating on various objects such as thoroughfares and vehicles. Limitations with the use of Google Street View present challenges in fully capturing the real experience of urban navigation. Future research involving mobile eye trackers in an actual physical environment could provide more comprehensive insights. Additional exploration of secondary objects might further contribute to our understanding of pedestrian navigation.

2. What are the patterns in duration and frequency of fixations seen across pedestrians, how do they differ among individuals, and how do these patterns reflect the navigational behaviors of the participants?

The examination of fixation behaviors and visual processing strategies revealed subtle and individualized patterns among pedestrians. Statistically significant differences were uncovered between participants who "Passed" and "Failed" the task in terms of overall fixation duration and fixation count, challenging conventional assumptions about fixation being indicative of better performance. This study highlights both shared and individual strategies in navigation and processing information, emphasizing the importance of efficiency in scanning. Overall, the research presents a view of navigation where rapid and efficient visual engagement might be more critical than previously thought, opening doors for further investigation.

3. What is the relationship between pedestrians' fixation patterns on urbanscape objects and their navigation performance in terms of successful reaching of the mapped destination?

The analysis of navigation performance and relationships between fixation patterns offered valuable insights. Distinct differences were identified between participants who successfully completed the task and those who did not. The findings affirm that closely following the optimal path leads to more successful navigation, while greater deviations are associated with failure. The study also identified correlations between deviation from the optimal path and various aspects of fixation behavior. These insights have profound implications for real-world applications, including the design of navigational tools and training programs. The results reflect the nuanced nature of navigation and open avenues for further exploration in various scenarios.

4. Based on the understanding of the key urbanscape objects that attract pedestrians' visual attention during self-navigation, what insights and recommendations can be provided for urban design and planning?

Addressing the research question, the current study reinforces the idea that buildings, especially their facades, play a vital role in shaping an individual's understanding of an urban landscape. However, it's crucial to recognize that nuances of facade distinction may require further exploration. Future research might shed more light on how these architectural elements influence navigation. Thoroughfares also emerged as significant points of fixation. The study suggests the potential value in incorporating unique navigational cues into pedestrian pathways, adding to the importance of pedestrian-friendly cities. The insights provided here are context-specific, and further research will be instrumental in forging comprehensive and nuanced guidelines for urban design and planning.

This research contributes an analysis that examines the complex relationship between urban architecture, and human navigation. It also revealing insights into human visual processing and navigation. The study acknowledges certain limitations, such as its reliance on a 2D environment, highlighting potential areas for further exploration that may enhance both theoretical understanding and practical application in the field. The findings lay a foundational understanding upon which future research can expand, potentially impacting urban planning in a broader context.

5.2. Future Directions and Limitations of the Research

This study presents several limitations that must be considered. First, the use of a 2D environment based on Google Street View does not entirely capture the complexity of navigating in a real 3D city. The results may be influenced by this difference in dimensionality. Future work could employ immersive 3D platforms or even consider mobile eye tracking in actual urban contexts for more realistic and robust findings.

Secondly, both the semantic segmentation model and the NLTK label classifications used in this study, though proficient in their tasks, have room for improvement. The segmentation model could benefit from enhanced object detail recognition, providing more granular classifications beyond broad categories like "building." Likewise, more refined NLTK label classifications could yield a richer understanding of the data.

The third limitation is related to the computational demands and downsampling to lower FPS. The slow processing speed of the semantic segmentation model not only restricted the sample size but might have obscured patterns in the data. Additionally, the cost of the hardware required for optimal performance could be a barrier for some researchers.

Future research directions should address these limitations. Adopting 3D environments or actual on-site mobile eye tracking, using more sophisticated semantic segmentation models capable of recognizing intricate object details, and improving NLTK label classifications would be promising avenues. Additionally, employing faster or more efficient models and leveraging higher frequency and precision data would enable studies with larger sample sizes and more accurate results.

These advancements would not only mitigate the current study's limitations but also contribute to a deeper understanding of the complex interplay between urban design and human navigation. By focusing on these areas, future research can build on this study's foundation, offering valuable insights and tools for both researchers and practitioners in the field.

LIST OF REFERENCES

6. List of References

Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: A synthesis and taxonomy. *Brain: A Journal of Neurology*, *122 (Pt 9)*, 1613–1628. https://doi.org/10.1093/brain/122.9.1613

Allen, G. L. (1999). Cognitive Abilities in the Service of Wayfinding: A Functional Approach. *The Professional Geographer*, *51*(4), 555–561. https://doi.org/10.1111/0033-0124.00192

Andersen, N. E., Dahmani, L., Konishi, K., & Bohbot, V. D. (2012). Eye tracking, strategies, and sex differences in virtual navigation. *Neurobiology of Learning and Memory*, *97*(1), 81–89. https://doi.org/10.1016/j.nlm.2011.09.007

Benson, A. (2003, February 1). *Spatial Disorientation—A Perspective*. https://www.semanticscholar.org/paper/Spatial-Disorientation-A-Perspective-Benson/6cb9956c0d3e91043af775f63ff1f5ac8a669dd4

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*.

Birsan, T., & Tiba, D. (2006). One Hundred Years Since the Introduction of the Set Distance by Dimitrie Pompeiu. In F. Ceragioli, A. Dontchev, H. Futura, K. Marti, & L. Pandolfi (Eds.), *System Modeling and Optimization* (pp. 35–39). Springer US. https://doi.org/10.1007/0-387-33006-2_4

Boada, I., Navazo, I., & Scopigno, R. (2001). Multiresolution volume visualization with a texture-based octree. *The Visual Computer*, *17*(3), 185–197. https://doi.org/10.1007/PL00013406

Bongiorno, C., Zhou, Y., Kryven, M., Theurel, D., Rizzo, A., Santi, P., Tenenbaum, J., & Ratti, C. (2021). Vector-based pedestrian navigation in cities. *Nature Computational Science*, *1*(10), Article 10. https://doi.org/10.1038/s43588-021-00130-y

Brunyé, T. T., Collier, Z. A., Cantelon, J., Holmes, A., Wood, M. D., Linkov, I., & Taylor, H. A. (2015). Strategies for Selecting Routes through Real-World Environments: Relative Topography, Initial Route Straightness, and Cardinal Direction. *PLOS ONE*, *10*(5), e0124404. https://doi.org/10.1371/journal.pone.0124404

Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology*. John Wiley & Sons.

Byrne, S. A., Maquiling, V., Reynolds, A. P. F., Polonio, L., Castner, N., & Kasneci, E. (2023). Exploring the Effects of Scanpath Feature Engineering for Supervised Image Classification Models. *Proceedings of the ACM on Human-Computer Interaction*, *7*(ETRA), 1–18. https://doi.org/10.1145/3591130

Cabanek, A., Zingoni de Baro, M. E., & Newman, P. (2020). Biophilic streets: A design framework for creating multiple urban benefits. *Sustainable Earth*, *3*(1), 7. https://doi.org/10.1186/s42055-020-00027-0

Caduff, D., & Timpf, S. (2008). On the assessment of landmark salience for human navigation. *Cognitive Processing*, *9*(4), 249–267. https://doi.org/10.1007/s10339-007-0199-2

Caldani, S., Isel, F., Septier, M., Acquaviva, E., Delorme, R., & Bucci, M. P. (2020). Impairment in Attention Focus During the Posner Cognitive Task in Children With ADHD: An Eye Tracker Study. *Frontiers in Pediatrics*, *8*. https://www.frontiersin.org/articles/10.3389/fped.2020.00484

Carbonell-Carrera, C., & Saorín, J. L. (2017). Geospatial Google Street View with Virtual Reality: A Motivational Approach for Spatial Training Education. *ISPRS International Journal of Geo-Information*, *6*(9), Article 9. https://doi.org/10.3390/ijgi6090261

Chan, E., Baumann, O., Bellgrove, M., & Mattingley, J. (2012). From Objects to Landmarks: The Function of Visual Location Information in Spatial Navigation. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00304

Chen, Jiaqi, Yang, Zeyu, & Zhang Li. (2023). *Semantic Segment Anything*. https://github.com/fudan-zvg/Semantic-Segment-Anything

Cheng, B., Luo, X., Mei, X., Chen, H., & Huang, J. (2022). A Systematic Review of Eye-Tracking Studies of Construction Safety. *Frontiers in Neuroscience*, *16*. https://www.frontiersin.org/articles/10.3389/fnins.2022.891725

Dalton, R. (2001). Spatial Navigation in Immersive Virtual Environments. *Conroy Dalton, R. (2001) Spatial Navigation in Immersive Virtual Environments. Doctoral Thesis, University of London.*

Davies, A., Vigo, M., Harper, S., & Jay, C. (2016). The visualisation of eye-tracking scanpaths: What can they tell us about how clinicians view electrocardiograms? *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)*, 79–83. https://doi.org/10.1109/ETVIS.2016.7851172

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive*, *16*, 409–458.

Dodsworth, C., Norman, L. J., & Thaler, L. (2020). Navigation and perception of spatial layout in virtual echo-acoustic space. *Cognition*, *197*, 104185. https://doi.org/10.1016/j.cognition.2020.104185

Dombeck, D. A., & Reiser, M. B. (2012). Real neuroscience in virtual worlds. *Current Opinion in Neurobiology*, *22*(1), 3–10.

Ekstrom, A. D. (2015). Why vision is important to how we navigate. *Hippocampus*, *25*(6), 731–735. https://doi.org/10.1002/hipo.22449

Enders, L., Smith, R., Gordon, S., Ries, A., & Touryan, J. (2021). *Identification of Target Objects from Gaze Behavior during a Virtual Navigation Task*. https://doi.org/10.1101/2021.03.30.437718

Fang, Z., Li, Q., & Shaw, S.-L. (2015). What about people in pedestrian navigation? *Geo-Spatial Information Science*, *18*(4), 135–150. https://doi.org/10.1080/10095020.2015.1126071

Farr, A. C., Kleinschmidt, T., Yarlagadda, P., & Mengersen, K. (2012). Wayfinding: A simple concept, a complex process. *Transport Reviews*, *32*(6), 715–743. https://doi.org/10.1080/01441647.2012.712555

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, *16*(3), 553–567. https://doi.org/10.1017/S0140525X00031575

Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(2), 195–215. https://doi.org/10.1037/0278-7393.31.2.195

Ford, P., Fisher, J., Paxman-Clarke, L., & Minichiello, M. (2020). Effective wayfinding adaptation in an older National Health Service hospital in the United Kingdom: Insights from mobile eye-tracking. *Design for Health*, *4*(1), 105–121. https://doi.org/10.1080/24735132.2020.1729000

Franke, C., & Schweikart, J. (2017). Investigation of Landmark-Based Pedestrian Navigation Processes with a Mobile Eye Tracking System. In G. Gartner & H. Huang (Eds.), *Progress in Location-Based Services 2016* (pp. 105–130). Springer International Publishing. https://doi.org/10.1007/978-3-319-47289-8_6

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Psychology Press.

Golledge. (1999). *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. JHU Press.

Golledge, R. G. (1993). Chapter 2 Geographical Perspectives on Spatial Cognition. In T. Gärling & R. G. Golledge (Eds.), *Advances in Psychology* (Vol. 96, pp. 16–46). North-Holland. https://doi.org/10.1016/S0166-4115(08)60038-2

Google LLC. (2022). *The Cathedral of Florence*. https://www.google.nl/maps/@43.7709145,11.2647744,3a,90y,298.06h,100.48t/data=!3m7!1e1!3m5!1si64eoceO1iNBs2hMS0wnsg!2e0!6shttps:%2F%2Fstreetviewpixels-pa.googleapis.com%2Fv1%2Fthumbnail%3Fpanoid%3Di64eoceO1iNBs2hMS0wnsg%26cb_client%3Dmaps_sv.tactile.gps%26w%3D203%26h%3D100%26yaw%3D262.01398%26pitch%3D0%26thumbfov%3D100!7i16384!8i8192?entry=ttu

Gresty, M. A., Golding, J. F., Le, H., & Nightingale, K. (2008). Cognitive Impairment by Spatial Disorientation. *Aviation, Space, and Environmental Medicine*, *79*(2), 105–111. https://doi.org/10.3357/ASEM.2143.2008

Grindinger, T. J., Murali, V. N., Tetreault, S., Duchowski, A. T., Birchfield, S. T., & Orero, P. (2011). Algorithm for Discriminating Aggregate Gaze Points: Comparison with Salient Regions-Of-Interest. In R. Koch & F. Huang (Eds.), *Computer Vision – ACCV 2010 Workshops* (pp. 390–399). Springer. https://doi.org/10.1007/978-3-642-22822-3_39

Gunzelmann, G., Anderson, J. R., & Douglass, S. (2004). Orientation Tasks with Multiple Views of Space: Strategies and Performance. *Spatial Cognition & Computation*, *4*(3), 207–253. https://doi.org/10.1207/s15427633scc0403_2

Hagler, G. S. W., Yelverton, T. L. B., Vedantham, R., Hansen, A. D. A., & Turner, J. R. (2011). Post-processing Method to Reduce Noise while Preserving High Time Resolution in Aethalometer Real-time Black Carbon Data. *Aerosol and Air Quality Research*, *11*(5), 539–546. https://doi.org/10.4209/aaqr.2011.05.0055

Hamid, S. N., Stankiewicz, B., & Hayhoe, M. (2010). Gaze patterns in navigation: Encoding information in large-scale environments. *Journal of Vision*, *10*(12), 28. https://doi.org/10.1167/10.12.28

Hejtmánek, L., Oravcová, I., Motýl, J., Horáček, J., & Fajnerová, I. (2018). Spatial knowledge impairment after GPS guided navigation: Eye-tracking study in a virtual town. *International Journal of Human-Computer Studies*, *116*, 15–24. https://doi.org/10.1016/j.ijhcs.2018.04.006

Hill, L., Townsend, J., Snider, J., Spence, R., Engler, A.-M., Moran, R., Hacker, S., & Chukoskie, L. (2020). *Distraction 'Hangover': Characterization of the Delayed Return to Baseline Driving Risk After Distracting Behaviors*. https://doi.org/10.7922/G2377706

Hofmaenner, D. A., Herling, A., Klinzing, S., Wegner, S., Lohmeyer, Q., Schuepbach, R. A., & Buehler, P. K. (2021). Use of eye tracking in analyzing distribution of visual attention among critical care nurses in daily professional life: An observational study. *Journal of Clinical Monitoring and Computing*, *35*(6), 1511–1518. https://doi.org/10.1007/s10877-020-00628-2

Hollander, J. B., Sussman, A., Lowitt, P., Angus, N., & Situ, M. (2021). Eye-tracking emulation software: A promising urban design tool. *Architectural Science Review*, *64*(4), 383–393. https://doi.org/10.1080/00038628.2021.1929055

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hutton, S. (2020, July 2). *Eye Tracking Terminology—Eye Movements*. Fast, Accurate, Reliable Eye Tracking. https://www.sr-research.com/eye-tracking-blog/background/eye-tracking-terminology-eye-movements/

Iftikhar, H., Shah, P., & Luximon, Y. (2021). Human wayfinding behaviour and metrics in complex environments: A systematic literature review. *Architectural Science Review*, *64*(5), 452–463. https://doi.org/10.1080/00038628.2020.1777386

Jamshidi, S., & Pati, D. (2021). A Narrative Review of Theories of Wayfinding Within the Interior Environment. *HERD: Health Environments Research & Design Journal*, *14*(1), 290–303. https://doi.org/10.1177/1937586720932276

Kiefer, P., Giannopoulos, I., & Raubal, M. (2014). Where Am I? Investigating Map Matching During Self-Localization With Mobile Eye Tracking in an Urban Environment. *Transactions in GIS*, *18*(5), 660–686. https://doi.org/10.1111/tgis.12067

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023, April 5). *Segment Anything*. ArXiv.Org. https://arxiv.org/abs/2304.02643v1

Lei, T.-C., Wu, S.-C., Chao, C.-W., & Lee, S.-H. (2016). Evaluating differences in spatial visual attention in wayfinding strategy when using 2D and 3D electronic maps. *GeoJournal*, *81*(2), 153–167. https://doi.org/10.1007/s10708-014-9605-3

Li, X., Wu, X.-Q., Yin, Z.-H., & Shen, J. (2017). THE INFLUENCE OF SPATIAL FAMILIARITY ON THE LANDMARK SALIENCE SENSIBILITY IN PEDESTRIAN NAVIGATION ENVIRONMENT. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2-W7*, 83–89. https://doi.org/10.5194/isprs-archives-XLII-2-W7-83-2017

Liao, H., Dong, W., Peng, C., & Liu, H. (2017). Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers. *Cartography and Geographic Information Science*, *44*(6), 474–490. https://doi.org/10.1080/15230406.2016.1174886

Lynch, K. (1964). *The Image of the City*. MIT Press.

Mahanama, B., Jayawardana, Y., Rengarajan, S., Jayawardena, G., Chukoskie, L., Snider, J., & Jayarathna, S. (2022). Eye Movement and Pupil Measures: A Review. *Frontiers in Computer Science*, *3*. https://www.frontiersin.org/articles/10.3389/fcomp.2021.733531

Meghanathan, R. N., van Leeuwen, C., & Nikolaev, A. R. (2015). Fixation duration surpasses pupil size as a measure of memory load in free viewing. *Frontiers in Human Neuroscience*, *8*. https://www.frontiersin.org/articles/10.3389/fnhum.2014.01063

Meilinger, T., Hölscher, C., Büchner, S. J., & Brösamle, M. (2007). How Much Information Do You Need? Schematic Maps in Wayfinding and Self Localisation. In T. Barkowsky, M. Knauff, G. Ligozat, & D. R. Montello (Eds.), *Spatial Cognition V Reasoning, Action, Interaction* (pp. 381–400). Springer. https://doi.org/10.1007/978-3-540-75666-8_22

Moffat, S. D., Hampson, E., & Hatzipantelis, M. (1998). Navigation in a "Virtual" Maze: Sex Differences and Correlation With Psychometric Measures of Spatial Ability in Humans. *Evolution and Human Behavior*, *19*(2), 73–87. https://doi.org/10.1016/S1090-5138(97)00104-9

Mohammadi Tahroodi, F., & Ujang, N. (2021). Engaging in social interaction: Relationships between the accessibility of path structure and intensity of passive social interaction in urban parks. *Archnet-IJAR: International Journal of Architectural Research*, *16*(1), 112–133. https://doi.org/10.1108/ARCH-04-2021-0100

Montello, D. R. (2005). Navigation. In A. Miyake & P. Shah (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 257–294). Cambridge University Press. https://doi.org/10.1017/CBO9780511610448.008

Moore, C. M., Yantis, S., & Vaughan, B. (1998). Object-Based Visual Selection: Evidence From Perceptual Completion. *Psychological Science*, *9*(2), 104–110. https://doi.org/10.1111/1467-9280.00019

Müller, T., Cotterell, R., Fraser, A., & Schütze, H. (2015). Joint Lemmatization and Morphological Tagging with Lemming. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2268–2274. https://doi.org/10.18653/v1/D15-1272

Netzel, R., Ohlhausen, B., Kurzhals, K., Woods, R., Burch, M., & Weiskopf, D. (2017). User performance and reading strategies for metro maps: An eye tracking study. *Spatial Cognition & Computation*, *17*(1–2), 39–64. https://doi.org/10.1080/13875868.2016.1226839

Oster, L. (2001). Using the think-aloud for reading instruction. *The Reading Teacher*, *55*, 64–69.

Peebles, D., Davies, C., & Mora, R. (2007). Effects of Geometry, Landmarks and Orientation Strategies in the 'Drop-Off' Orientation Task. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers (Eds.), *Spatial Information Theory* (pp. 390–405). Springer. https://doi.org/10.1007/978-3-540-74788-8_24

Rai, A., & Borah, S. (2021). Study of Various Methods for Tokenization. In J. K. Mandal, S. Mukhopadhyay, & A. Roy (Eds.), *Applications of Internet of Things* (pp. 193–200). Springer. https://doi.org/10.1007/978-981-15-6198-6_18

Rajaraman, A., & Ullman, J. D. (Eds.). (2011). Data Mining. In *Mining of Massive Datasets* (pp. 1–17). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9781139058452.002

Rehrl, K., Häusler, E., & Leitinger, S. (2010). Comparing the effectiveness of GPS-enhanced voice guidance for pedestrians with metric-and landmark-based instruction sets. *Geographic Information Science: 6th International Conference, GIScience 2010, Zurich, Switzerland, September 14-17, 2010. Proceedings 6*, 189–203.

Rey, B., & Alcaiz, M. (2010). *Research in Neuroscience and Virtual Reality*. https://doi.org/10.5772/13198

Robinson, D. A. (1968). The oculomotor control system: A review. *Proceedings of the IEEE*, *56*(6), 1032–1049. https://doi.org/10.1109/PROC.1968.6455

Romanes, G. J., & Darwin, C. (1884). *Mental evolution in animals, with a posthumous essay on instinct* (p. 411). D Appleton & Company. https://doi.org/10.1037/12803-000

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1997). Navigating buildings in "desk-top" virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology: Applied*, *3*(2), 143–159. https://doi.org/10.1037/1076-898X.3.2.143

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA '00*, 71–78. https://doi.org/10.1145/355017.355028

Sandstrom, N. J., Kaufman, J., & A. Huettel, S. (1998). Males and females use different distal cues in a virtual environment navigation task1Published on the World Wide Web on 27 January 1998.1. *Cognitive Brain Research*, *6*(4), 351–360. https://doi.org/10.1016/S0926-6410(98)00002-0

Schwedes, C., & Wentura, D. (2016). Through the eyes to memory: Fixation durations as an early indirect index of concealed knowledge. *Memory & Cognition*, *44*(8), 1244–1258. https://doi.org/10.3758/s13421-016-0630-y

Sezgin, M., & Sankur, B. lent. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, *13*(1), 146–168.

Shelton, A. L., Marchette, S. A., & Furman, A. J. (2013). Chapter Six—A Mechanistic Approach to Individual Differences in Spatial Learning, Memory, and Navigation. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 59, pp. 223–259). Academic Press. https://doi.org/10.1016/B978-0-12-407187-2.00006-X

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 4. https://doi.org/10.1167/7.14.4

Trapsilo, P. (2016). A Think-Aloud Protocols as a Cognitive Strategy to Increase Students' Writing Narrative Skill at Efl Classroom. *PREMISE JOURNAL:ISSN Online: 2442-482x, ISSN Printed:*

2089-3345.

https://www.academia.edu/62491769/A_Think_Aloud_Protocols_as_a_Cognitive_Strategy_to_Increase_Students_Writing_Narrative_Skill_at_Efl_Classroom

van der Ham, I. J. M., Faber, A. M. E., Venselaar, M., van Kreveld, M. J., & Löffler, M. (2015). Ecological validity of virtual environments to assess human navigation ability. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00637

Viaene, P., Ooms, K., Vansteenkiste, P., Lenoir, M., & De Maeyer, P. (2014). The Use of Eye Tracking in Search of Indoor Landmarks. *ET4S@ GIScience*, 52–56.

Waller, D., Beall, A. C., & Loomis, J. M. (2004). Using virtual environments to assess directional knowledge. *Journal of Environmental Psychology*, *24*(1), 105–116. https://doi.org/10.1016/S0272-4944(03)00051-3

Ware, C., & Mikaelian, H. H. (1986). An evaluation of an eye tracker as a device for computer input2. *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, 183–188. https://doi.org/10.1145/29933.275627

Wiener, J. M., Büchner, S. J., & Hölscher, C. (2009). Taxonomy of Human Wayfinding Tasks: A Knowledge-Based Approach. *Spatial Cognition & Computation*, *9*(2), 152–165. https://doi.org/10.1080/13875860902906496

Winkler, S., & Subramanian, R. (2013). *Overview of Eye tracking Datasets*. 212–217. https://doi.org/10.1109/QoMEX.2013.6603239

Wolcott, M. D., & Lobczowski, N. G. (2021). Using cognitive interviews and think-aloud protocols to understand thought processes. *Currents in Pharmacy Teaching and Learning*, *13*(2), 181–188. https://doi.org/10.1016/j.cptl.2020.09.005

Yesiltepe, D., Conroy Dalton, R., & Ozbil Torun, A. (2021). Landmarks in wayfinding: A review of the existing literature. *Cognitive Processing*, *22*(3), 369–410. https://doi.org/10.1007/s10339-021-01012-x

Yoder, R. M., Clark, B. J., & Taube, J. S. (2011). Origins of landmark encoding in the brain. *Trends in Neurosciences*, *34*(11), 561–571. https://doi.org/10.1016/j.tins.2011.08.004

Yue, Z., Zhong, Y., & Cui, Z. (2022). Respondent Dynamic Attention to Streetscape Composition in Nanjing, China. *Sustainability*, *14*(22), Article 22. https://doi.org/10.3390/su142215209

Yuxin Wu & Alexander Kirillov. (2019). *Detectron2*. https://github.com/facebookresearch/detectron2

Zemblys, R., Niehorster, D. C., Komogortsev, O., & Holmqvist, K. (2018). Using machine learning to detect events in eye-tracking data. *Behavior Research Methods*, *50*(1), 160–181. https://doi.org/10.3758/s13428-017-0860-3

APPENDIX

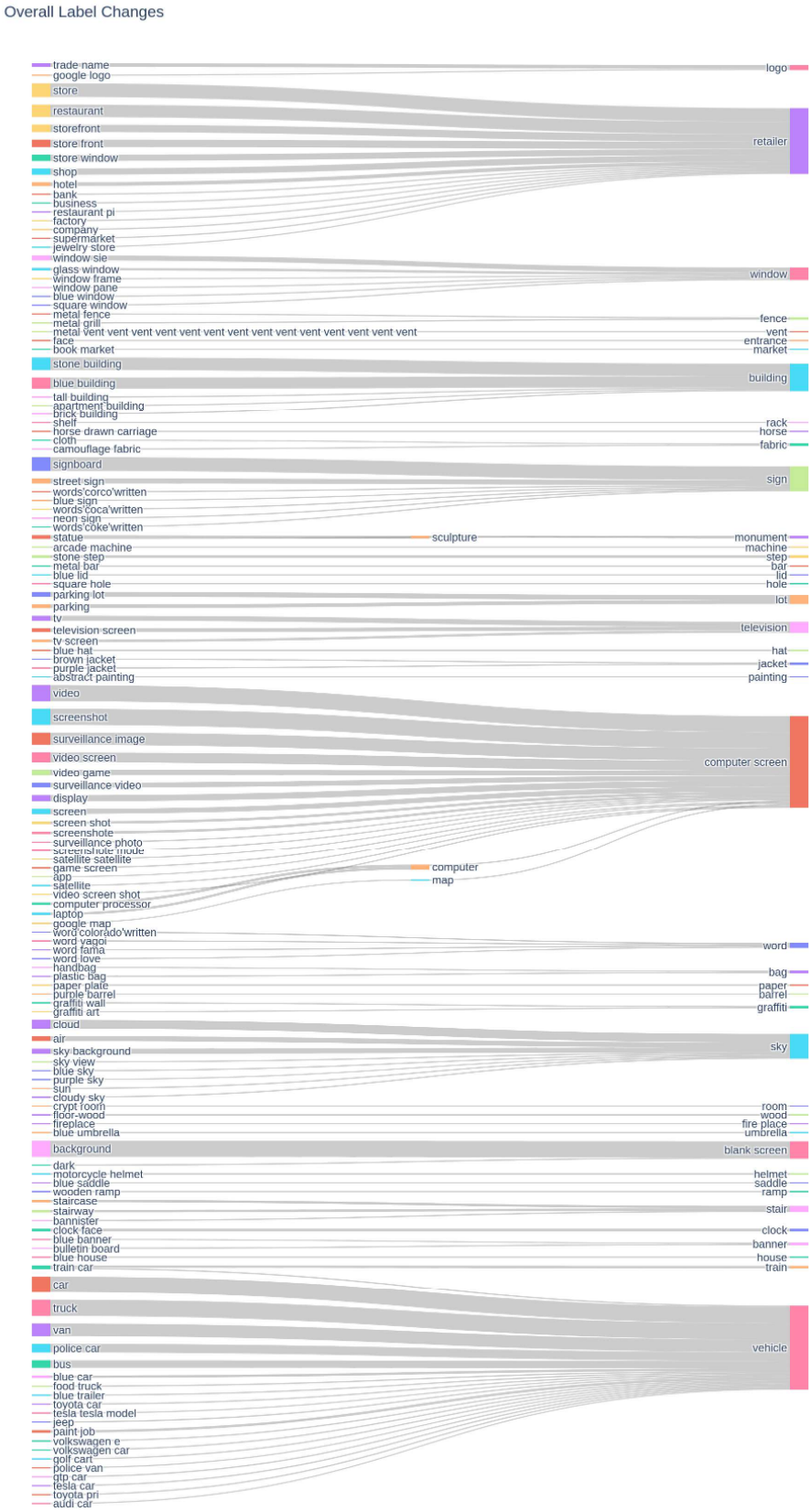
Annex 1: Eye-Tracking Study on Navigation in an Unfamiliar Urban Environment Protocol

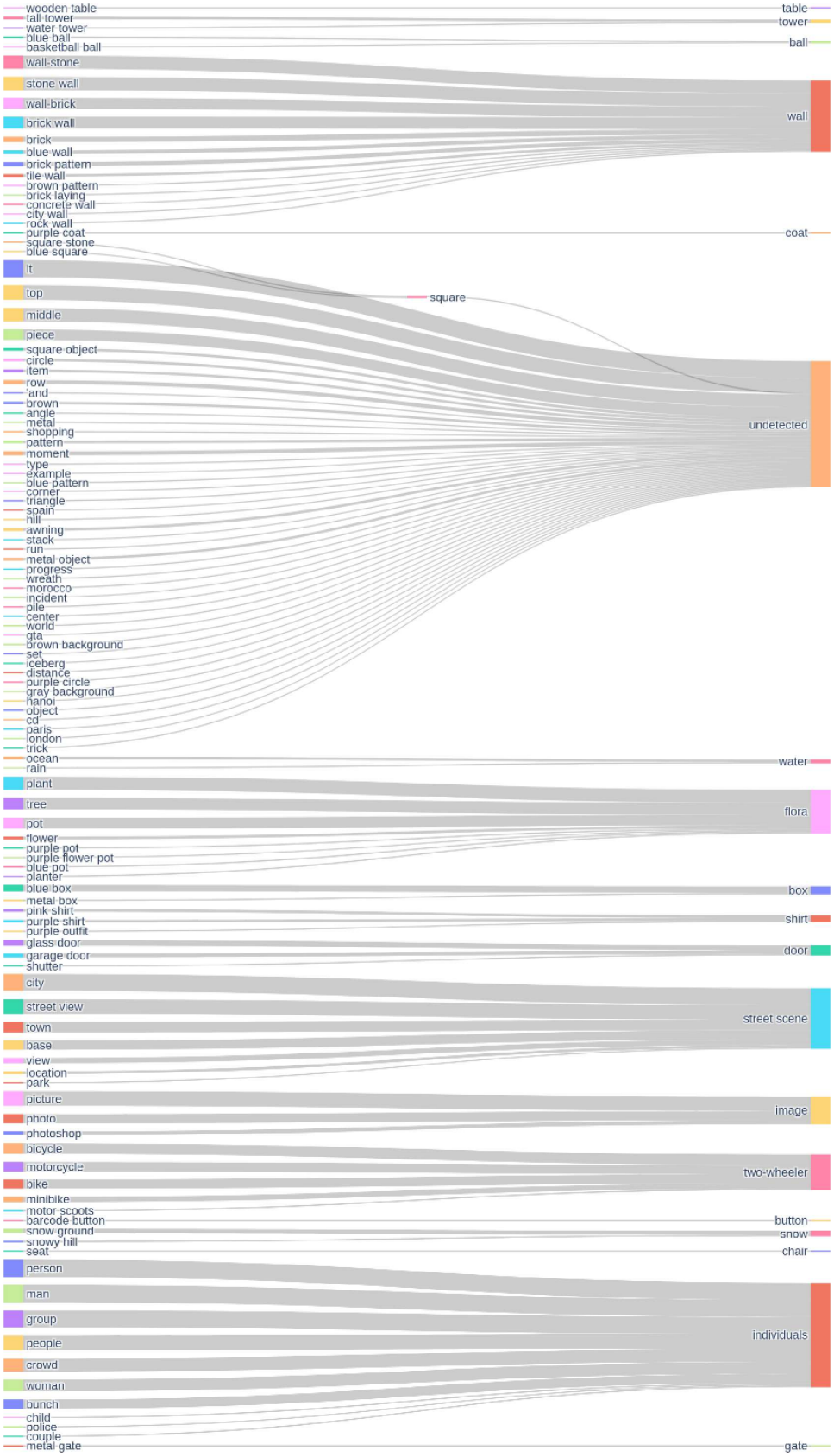
Protocol	*Eye-Tracking Study on Navigation in an Unfamiliar Urban Environment*
Objective	The objective of this study is to understand how individuals navigate in an unfamiliar urban environment using eye-tracking technology.
Participants	University students will be recruited on campus to participate in the study.
Materials	Computer with internet access API of Google Street View (Local server) Eye-tracking device (Tobii Fusion pro) Microphone Mouse and Keyboard
Introduction	Participants will be greeted by the researcher. An overview of the study's objective will be provided, emphasizing the focus on understanding navigation in unfamiliar urban environments using eye-tracking technology. Participants will be introduced to the eye-tracking technology and its role in monitoring eye movements during the navigation task.
Task	Participants will be presented with an environment created using the Google Street View API, with the bottom left corner map removed. The main task objective will be clearly communicated: to navigate from Point A, situated in front of Basilica di San Lorenzo, to Point B, located in front of Piazza della Signoria. The navigation task will conclude when the participant detects the third landmark and positions themselves in the plaza in front of it.
Introductory Video	A video will be shown to enhance participants' familiarity with the study area, featuring a 2D map of Florence and marking the start and end points. Three landmarks will be introduced during the video to assist participants in navigation.
Landmark Descriptions	Visual descriptions of each landmark will be provided, highlighting their distinctive characteristics. **Basilica di San Lorenzo:** • *Medium dome visible from various angles.* • *Brick walls surrounding the building.* • *A long wall attached to the building.* **Florence Cathedral:** • *White marble building with distinctive features.* • *A prominent dome that can be seen from various locations.* • *A large adjacent tower.* **Piazza della Signoria:**

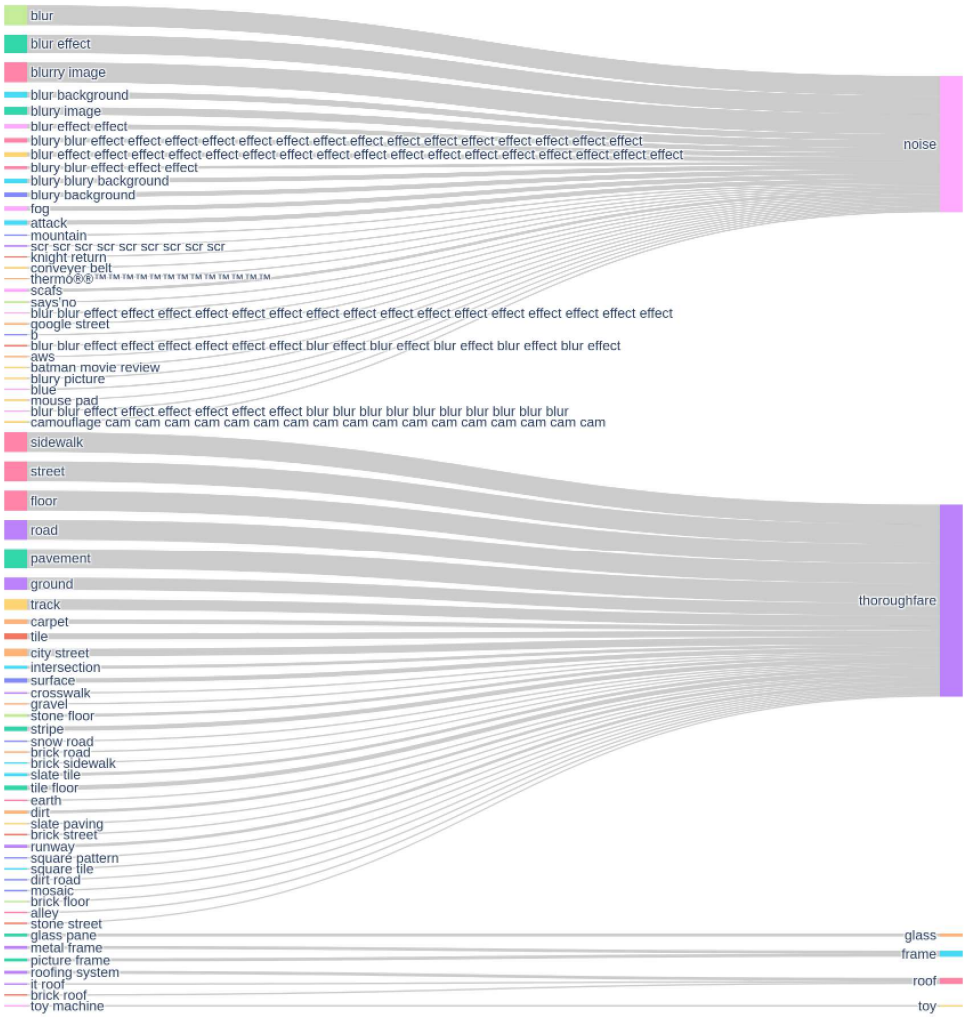
	• *A tower attached to a building in the vicinity.* • *Brick exterior around the plaza.* • *A spacious plaza serving as the final point of the navigation task.*
Navigation Instructions	Participants will use a mouse or keyboard to control movement within the virtual environment. Alternative methods for movement (arrow keys or arrow shapes on the pavement) will be advised in case images do not function optimally during navigation. Participants will be made aware of possible image rotations between each image and encouraged to adapt their movements to maintain accurate spatial orientation. The use of the mouse scroll function will be discouraged to avoid potential errors in research data.
Calibration and Eye-Tracking Setup	Participants will be seated comfortably in front of the eye tracker to minimize unnecessary movement. The "Tobii Pro Eye Tracker Manager" software will be used to assess the eye tracker's detection of participants' eyes and ensure optimal positioning. A calibration process will be conducted, where participants follow a moving dot on the screen with their eyes, establishing accurate and personalized gaze mapping.
Think-Aloud Analysis	Participants will engage in a "think-aloud" process, articulating their thoughts, decisions, and strategies aloud during navigation. The researcher may prompt participants with relevant questions if continuous verbalization is lacking.
Think-Aloud Prompts	During the navigation task, if a participant forgets to maintain the think-aloud verbalization, the researcher will use the following prompts to encourage participants to express their thoughts explicitly: • "What are you thinking right now?" • "Where do you want to go?" • "What do you think of this place?" • "Do you feel you're on the right track?" • "Do you know where you are going?" • "Do you feel lost?" The prompts will be used to facilitate the continuous expression of participants' thoughts and ensure a consistent stream of verbalization throughout the navigation task.
Duration	Participants will be informed that the navigational task has no predetermined time limit. However, the researcher may conclude the task if certain conditions are met, such as significant distance from the endpoint, task duration exceeding 20 minutes, or participant feeling completely lost.
Instructions and Support	Participants will have the opportunity to ask questions and seek clarification before beginning the task. Necessary instructions and support will be provided as needed.
Confidentiality and Data Usage	Participation in the study will be optional, and data will be treated confidentially for research purposes. Pseudonyms will be used instead of personal information to protect participants'

	privacy. Data collected will include demographic information, eye tracking data, and audio recordings during the think-aloud process. Participants will be asked to provide informed consent to participate in the survey.
Post-Survey Results Sharing	After completing the survey, participants will be shown eye tracking data and given a visual representation of their navigation. Participants will receive a small token of appreciation (e.g., chocolate) for their involvement.
Conclusion	Upon the completion of the study, the researchers will analyze the eye tracking data to gain insights into how individuals navigate in an unfamiliar urban environment. The findings will be shared with relevant stakeholders to contribute to the understanding of human navigation and eye movement patterns in such environments.

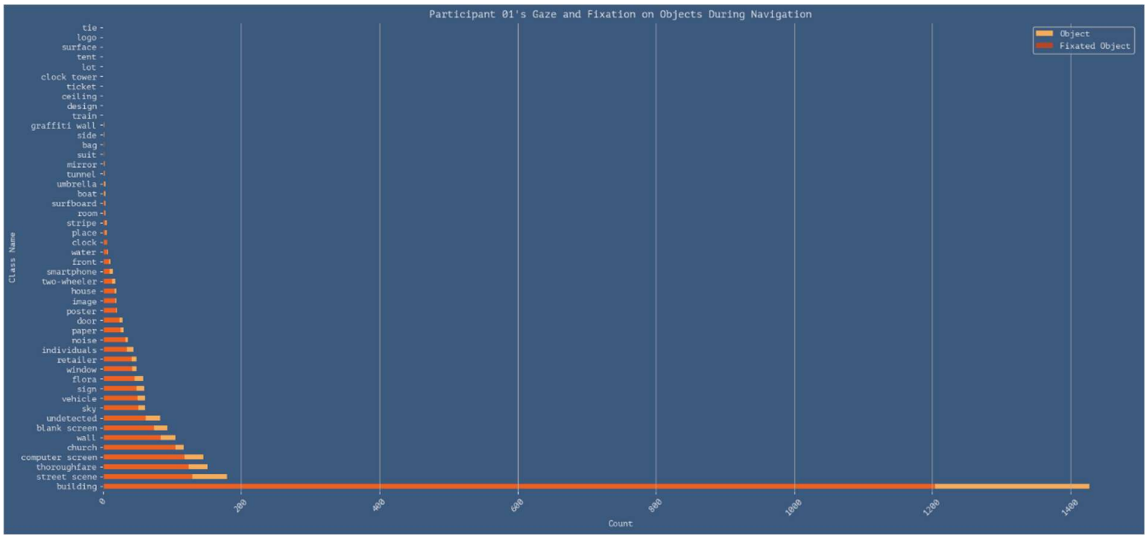
Annex 2: Overal Label Change

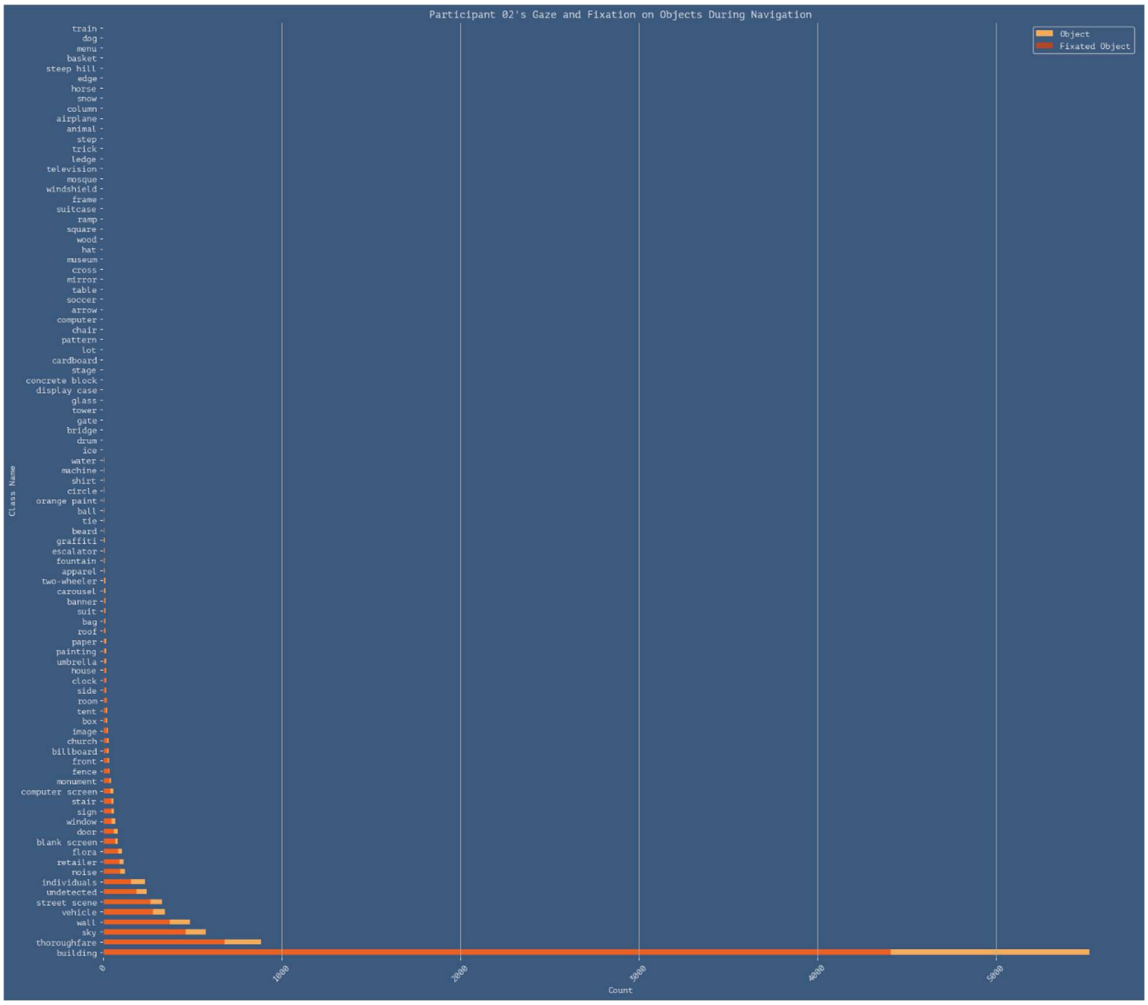


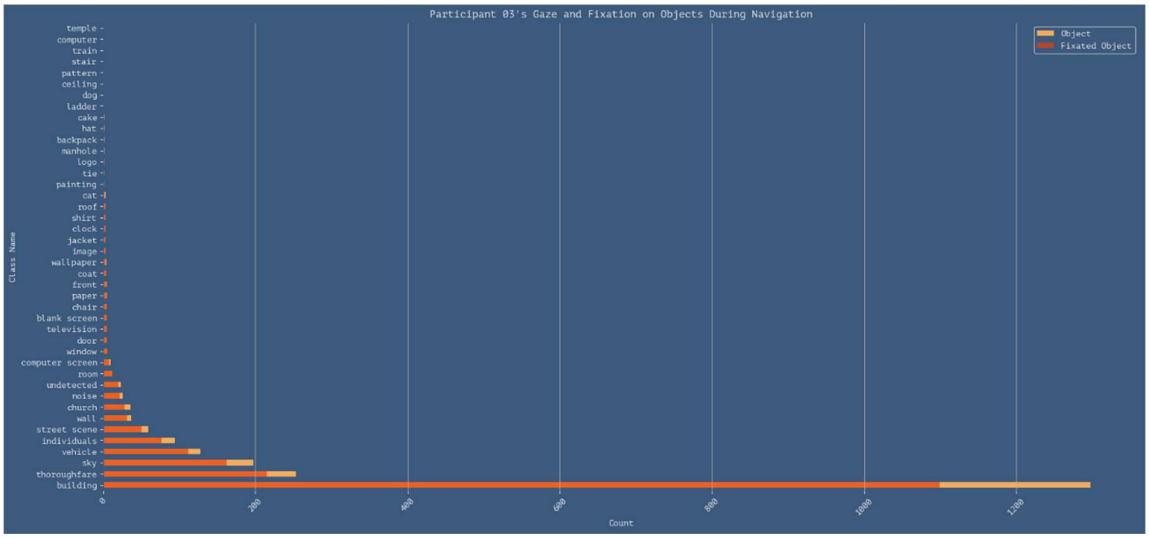

wooden table
tall tower
water tower
blue ball
basketball ball
wall:stone
stone wall
wall:brick
brick wall
brick
blue wall
brick pattern
tile wall
brown pattern
brick laying
concrete wall
city wall
rock wall
purple coat
square stone
blue square
it
top
middle
piece
square object
circle
item
row
'and
brown
angle
metal
shopping
pattern
moment
type
example
blue pattern
corner
triangle
spain
hill
awning
stack
run
metal object
progress
wreath
morocco
incident
pile
center
world
gta
brown background
set
iceberg
distance
purple circle
gray background
hanoi
object
cd
paris
london
trick
ocean
rain
plant
tree
pot
flower
purple pot
purple flower pot
blue pot
planter
blue box
metal box
pink shirt
purple shirt
purple outfit
glass door
garage door
shutter
city
street view
town
base
view
location
park
picture
photo
photoshop
bicycle
motorcycle
bike
minibike
motor scoots
barcode button
snow ground
snowy hill
seat
person
man
group
people
crowd
woman
bunch
child
police
couple
metal gate
square
table
tower
ball
wall
coat
undetected
water
flora
box
shirt
door
street scene
image
two-wheeler
button
snow
chair
individuals
gate

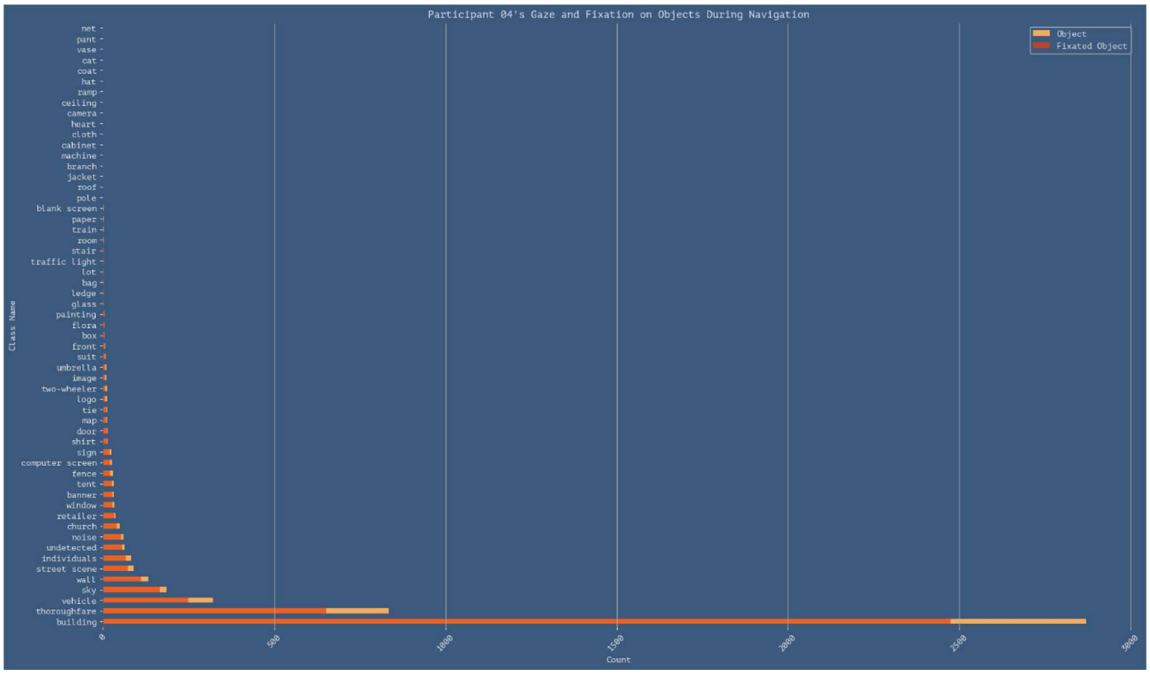

blur
blur effect
blurry image
blur background
blury image
blur effect effect
blury blur effect effect effect effect effect effect effect effect effect effect effect effect effect effect effect
blur effect effect effect effect effect effect effect effect effect effect effect effect effect effect effect effect
blury blur effect effect effect
blury blury background
blury background
fog
attack
mountain
scr scr scr scr scr scr scr scr scr scr
knight return
conveyer belt
thermo®™™™™™™™™™™™™™™™™
scafs
says'no
blur blur effect effect effect effect effect effect effect effect effect effect effect effect effect effect effect effect
google street
b
blur blur effect effect effect effect effect effect blur effect blur effect blur effect blur effect blur effect
aws
batman movie review
blury picture
blue
mouse pad
blur blur effect effect effect effect effect effect blur blur blur blur blur blur blur blur blur blur
camouflage cam cam cam cam cam cam cam cam cam cam cam cam cam cam cam cam cam
sidewalk
street
floor
road
pavement
ground
track
carpet
tile
city street
intersection
surface
crosswalk
gravel
stone floor
stripe
snow road
brick road
brick sidewalk
slate tile
tile floor
earth
dirt
slate paving
brick street
runway
square pattern
square tile
dirt road
mosaic
brick floor
alley
stone street
glass pane
metal frame
picture frame
roofing system
it roof
brick roof
toy machine
noise
thoroughfare
glass
frame
roof
toy

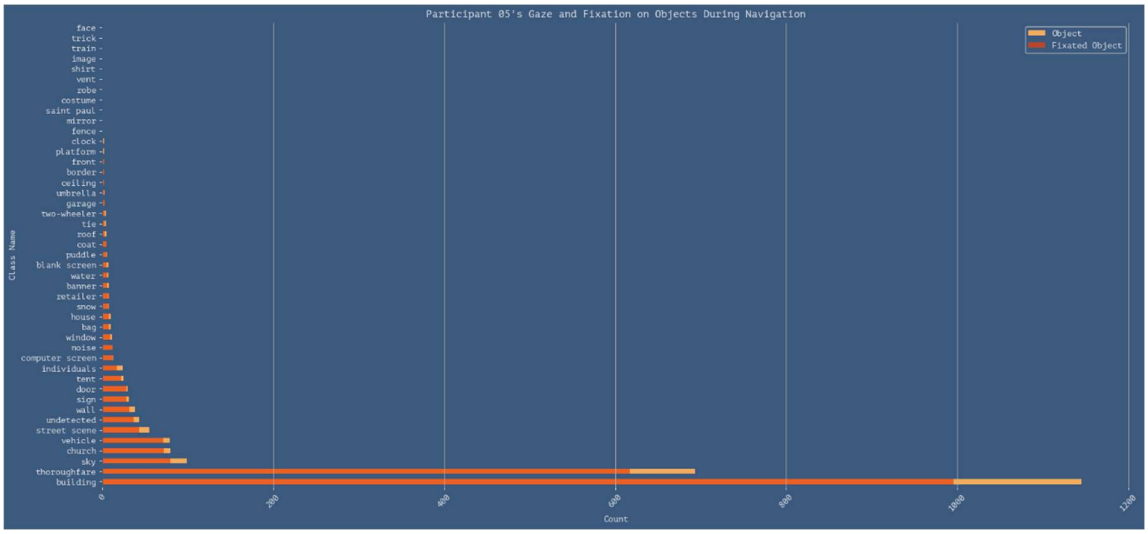
Annex 3: Figures related to Participants Gaze and Fixation on Objects During Navigation.

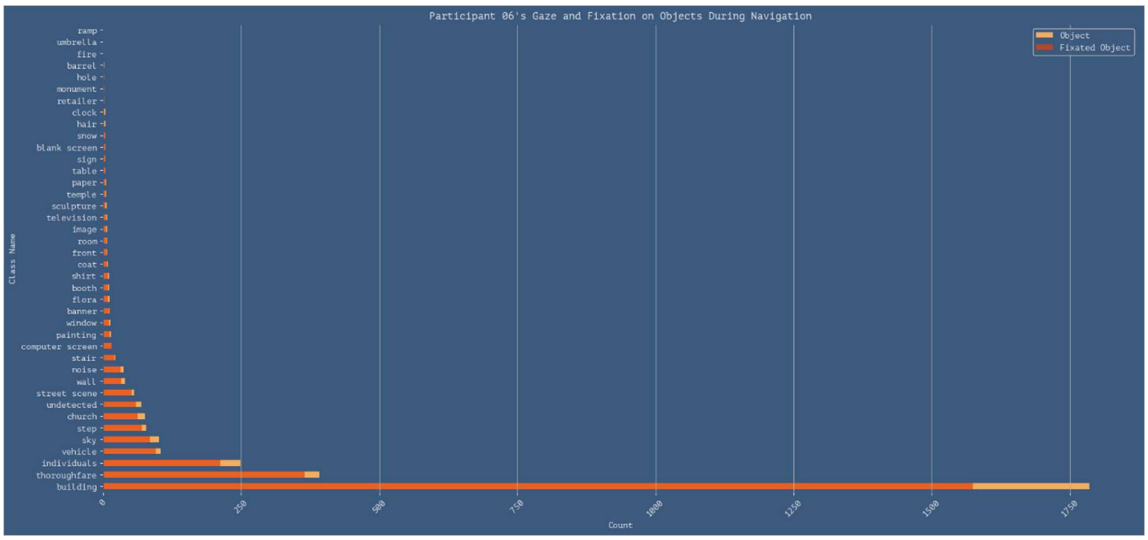


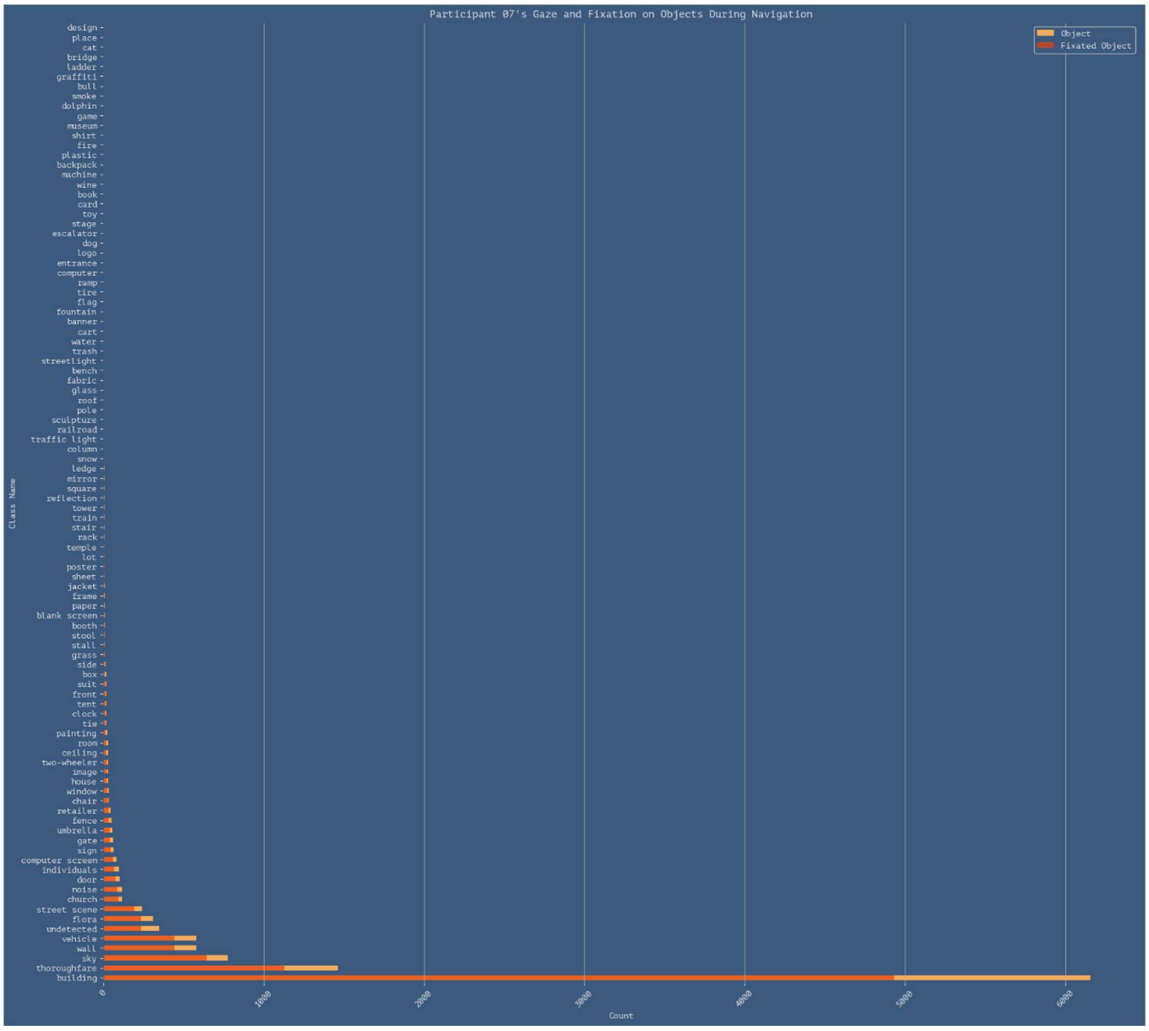


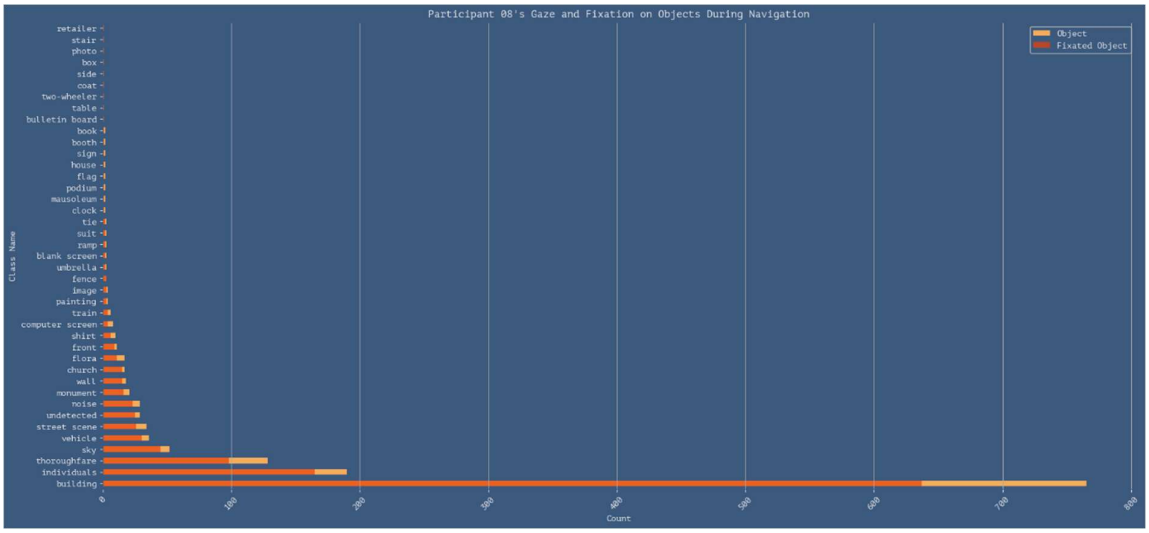

Participant 03's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
Class Name
temple
computer
train
stair
pattern
ceiling
dog
ladder
cake
hat
backpack
manhole
logo
tie
painting
cat
roof
shirt
clock
jacket
image
wallpaper
coat
front
paper
chair
blank screen
television
door
window
computer screen
room
undetected
noise
church
wall
street scene
individuals
vehicle
sky
thoroughfare
building
0
200
400
600
800
1000
1200
Count

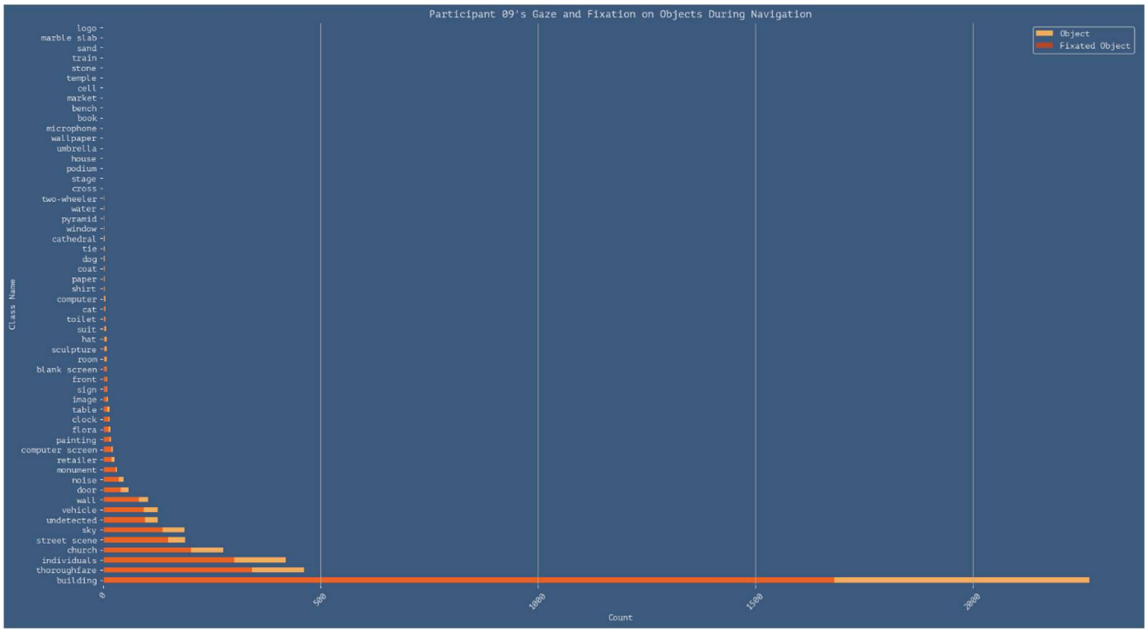

Participant 04's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
Class Name
net
pant
vase
cat
coat
hat
ramp
ceiling
camera
heart
cloth
cabinet
machine
branch
jacket
roof
pole
blank screen
paper
train
room
stair
traffic light
lot
bag
ledge
glass
painting
flora
box
front
suit
umbrella
image
two-wheeler
logo
tie
map
door
shirt
sign
computer screen
fence
tent
banner
window
retailer
church
noise
undetected
individuals
street scene
wall
sky
vehicle
thoroughfare
building
0
500
1000
1500
2000
2500
3000
Count

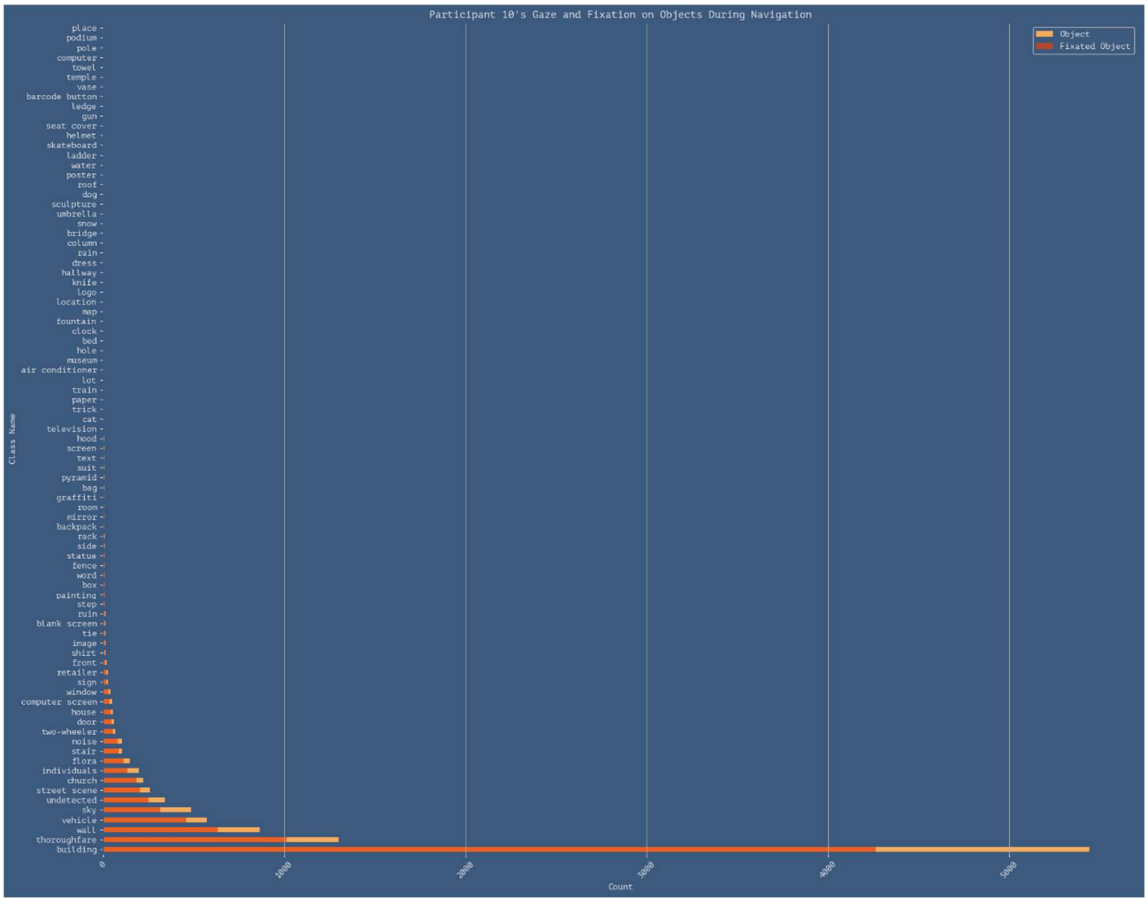

Participant 05's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
face
trick
train
image
shirt
vent
robe
costume
saint paul
mirror
fence
clock
platform
front
border
ceiling
umbrella
garage
two-wheeler
tie
roof
coat
puddle
blank screen
water
banner
retailer
snow
house
bag
window
noise
computer screen
individuals
tent
door
sign
wall
undetected
street scene
vehicle
church
sky
thoroughfare
building
Class Name
0
200
400
600
800
1000
1200
Count

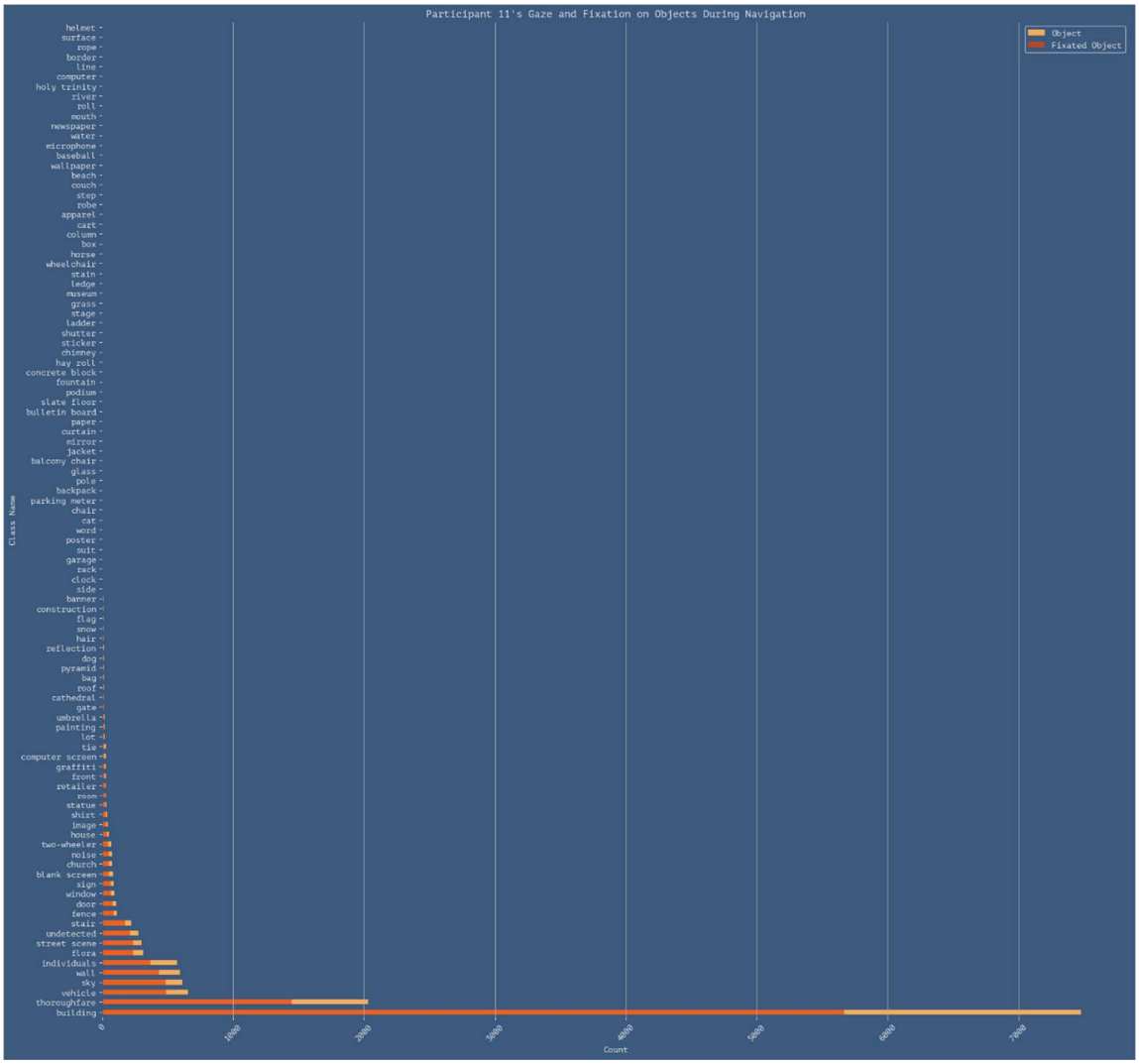

Participant 06's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
ramp
umbrella
fire
barrel
hole
monument
retailer
clock
hair
snow
blank screen
sign
table
paper
temple
sculpture
television
image
room
front
coat
shirt
booth
flora
banner
window
painting
computer screen
stair
noise
wall
street scene
undetected
church
step
sky
vehicle
individuals
thoroughfare
building
Class Name
0
250
500
750
1000
1250
1500
1750
Count

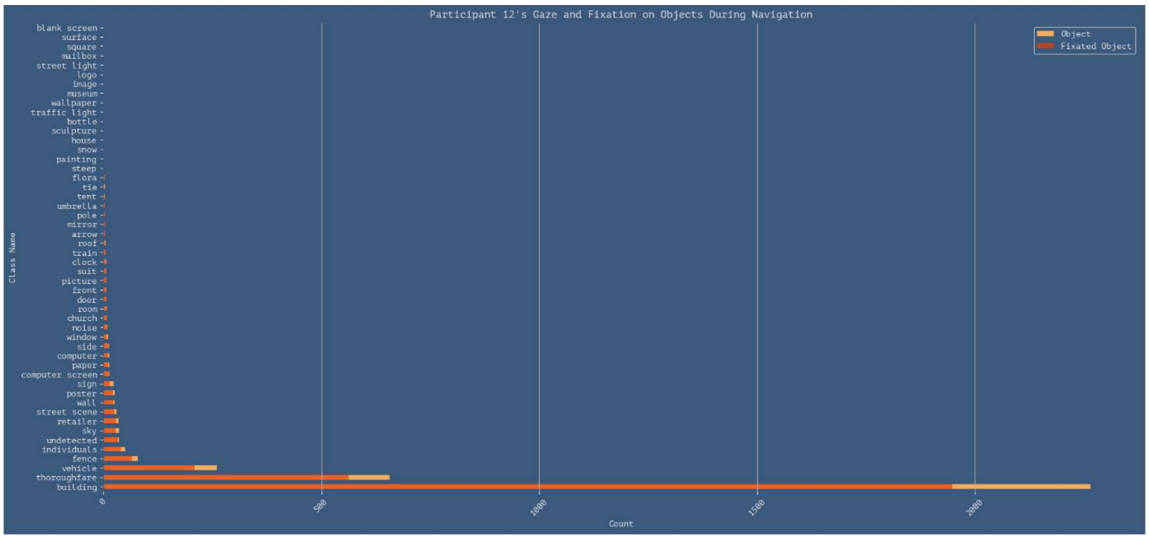

Participant 07's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
design
place
cat
bridge
ladder
graffiti
bull
smoke
dolphin
game
museum
shirt
fire
plastic
backpack
machine
wine
book
card
toy
stage
escalator
dog
logo
entrance
computer
ramp
tire
flag
fountain
banner
cart
water
trash
streetlight
bench
fabric
glass
roof
pole
sculpture
railroad
traffic light
column
snow
ledge
mirror
square
reflection
tower
train
stair
rack
temple
lot
poster
sheet
jacket
frame
paper
blank screen
booth
stool
stall
grass
side
box
suit
front
tent
clock
tie
painting
room
ceiling
two-wheeler
image
house
window
chair
retailer
fence
umbrella
gate
sign
computer screen
individuals
door
noise
church
street scene
flora
undetected
vehicle
wall
sky
thoroughfare
building
Class Name
0
1000
2000
3000
4000
5000
6000
Count

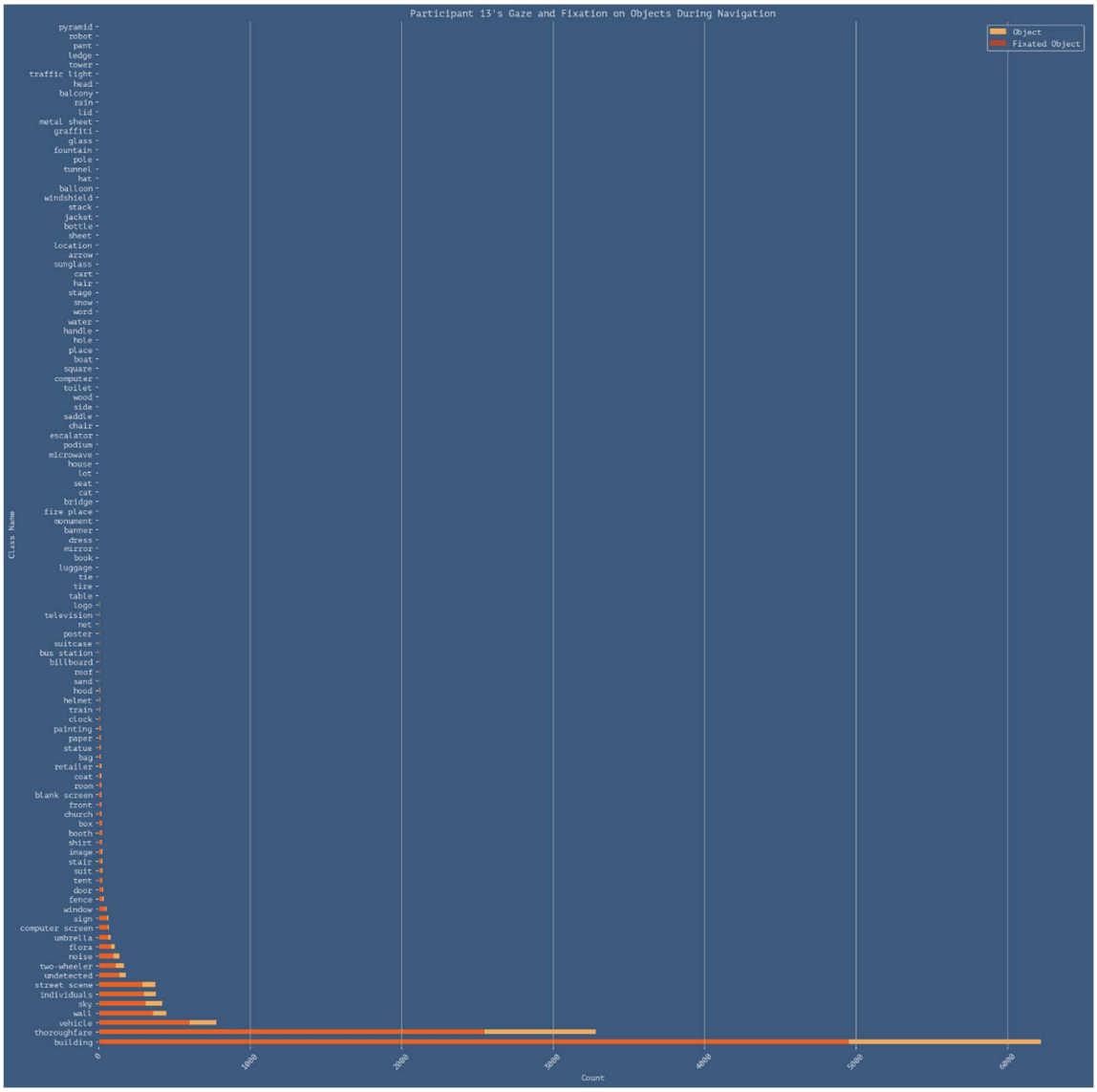

Participant 08's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
retailer
stair
photo
box
side
coat
two-wheeler
table
bulletin board
book
booth
sign
house
flag
podium
mausoleum
clock
tie
suit
ramp
blank screen
umbrella
fence
image
painting
train
computer screen
shirt
front
flora
church
wall
monument
noise
undetected
street scene
vehicle
sky
thoroughfare
individuals
building
Class Name
0
100
200
300
400
500
600
700
800
Count


Participant 09's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
Class Name
logo
marble slab
sand
train
stone
temple
cell
market
bench
book
microphone
wallpaper
umbrella
house
podium
stage
cross
two-wheeler
water
pyramid
window
cathedral
tie
dog
coat
paper
shirt
computer
cat
toilet
suit
hat
sculpture
room
blank screen
front
sign
image
table
clock
flora
painting
computer screen
retailer
monument
noise
door
wall
vehicle
undetected
sky
street scene
church
individuals
thoroughfare
building
0
500
1000
1500
2000
Count

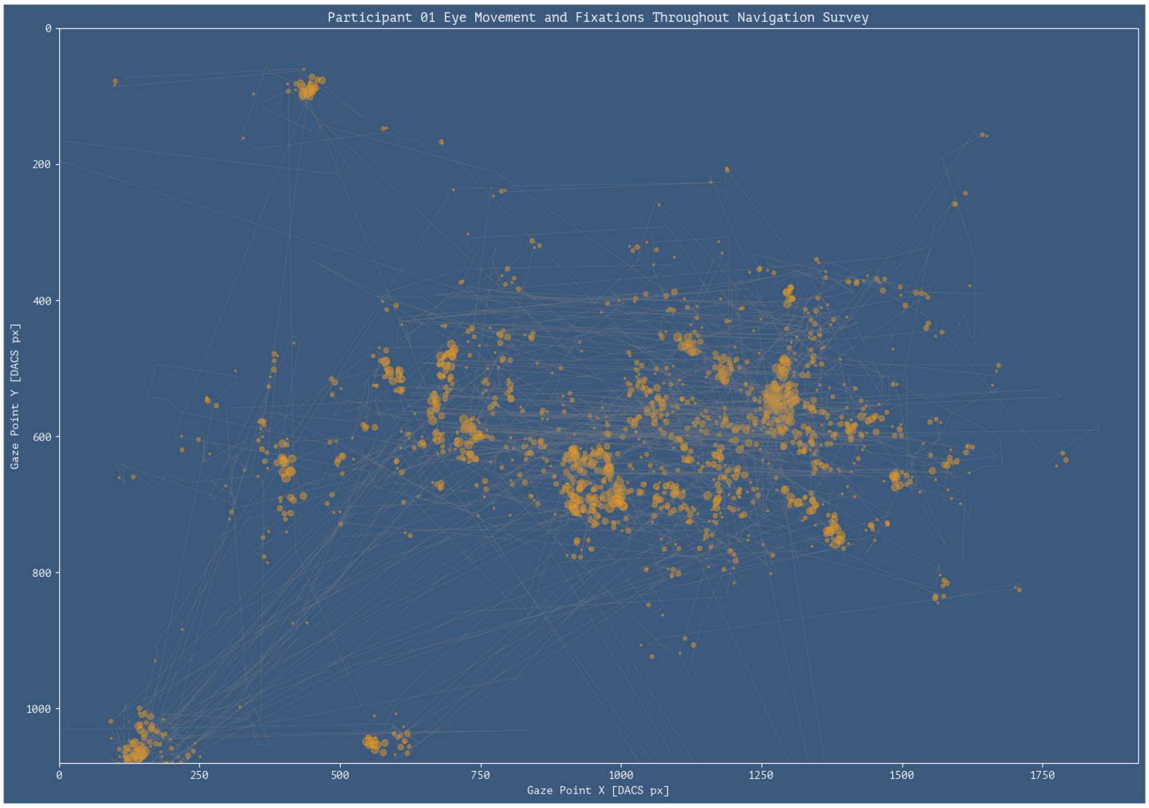

Participant 10's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
Class Name
place
podium
pole
computer
towel
temple
vase
barcode button
ledge
gun
seat cover
helmet
skateboard
ladder
water
poster
roof
dog
sculpture
umbrella
snow
bridge
column
rain
dress
hallway
knife
logo
location
map
fountain
clock
bed
hole
museum
air conditioner
lot
train
paper
trick
cat
television
hood
screen
text
suit
pyramid
bag
graffiti
room
mirror
backpack
rack
side
statue
fence
word
box
painting
step
ruin
blank screen
tie
image
shirt
front
retailer
sign
window
computer screen
house
door
two-wheeler
noise
stair
flora
individuals
church
street scene
undetected
sky
vehicle
wall
thoroughfare
building
0
1000
2000
3000
4000
5000
Count

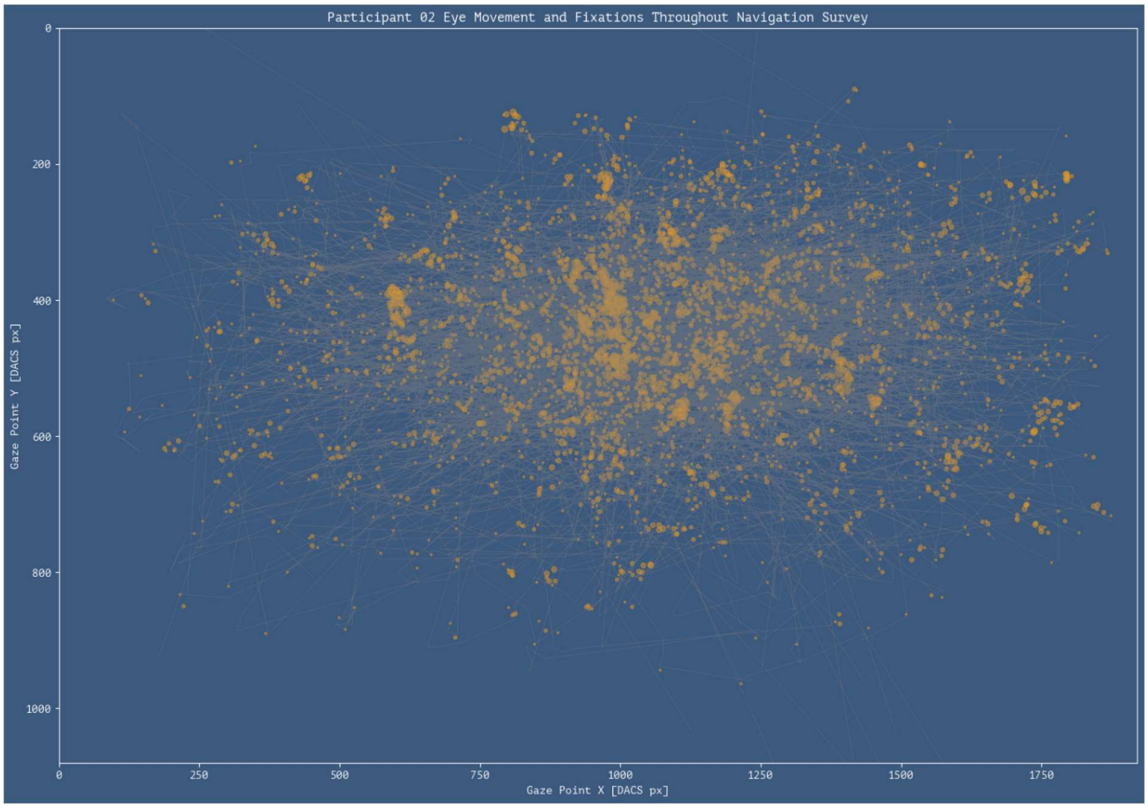

Participant 11's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
helmet
surface
rope
border
line
computer
holy trinity
river
roll
mouth
newspaper
water
microphone
baseball
wallpaper
beach
couch
step
robe
apparel
cart
column
box
horse
wheelchair
stain
ledge
museum
grass
stage
ladder
shutter
sticker
chimney
hay roll
concrete block
fountain
podium
slate floor
bulletin board
paper
curtain
mirror
jacket
balcony chair
glass
pole
backpack
parking meter
chair
cat
word
poster
suit
garage
rack
clock
side
banner
construction
flag
snow
hair
reflection
dog
pyramid
bag
roof
cathedral
gate
umbrella
painting
lot
tie
computer screen
graffiti
front
retailer
room
statue
shirt
image
house
two-wheeler
noise
church
blank screen
sign
window
door
fence
stair
undetected
street scene
flora
individuals
wall
sky
vehicle
thoroughfare
building
Class Name
0
1000
2000
3000
4000
5000
6000
7000
Count

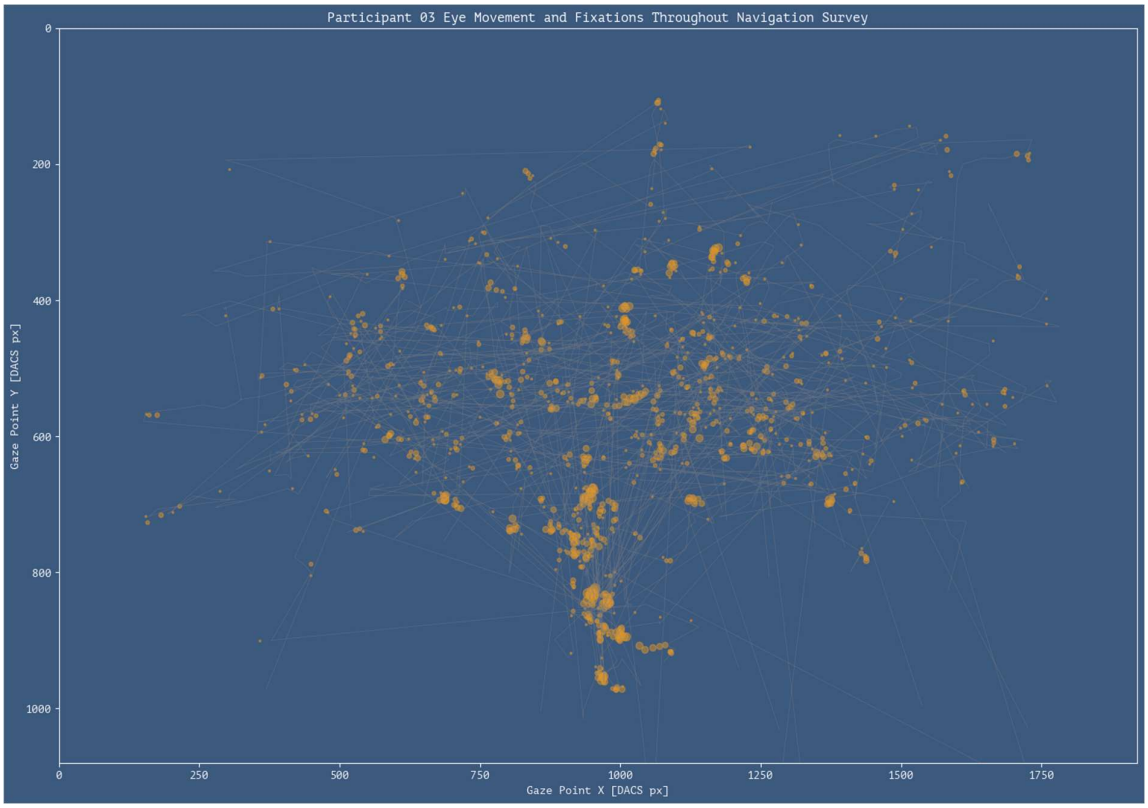

Participant 12's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
blank screen
surface
square
mailbox
street light
logo
image
museum
wallpaper
traffic light
bottle
sculpture
house
snow
painting
steep
flora
tie
tent
umbrella
pole
mirror
arrow
roof
train
clock
suit
picture
front
door
room
church
noise
window
side
computer
paper
computer screen
sign
poster
wall
street scene
retailer
sky
undetected
individuals
fence
vehicle
thoroughfare
building
Class Name
0
500
1000
1500
2000
Count

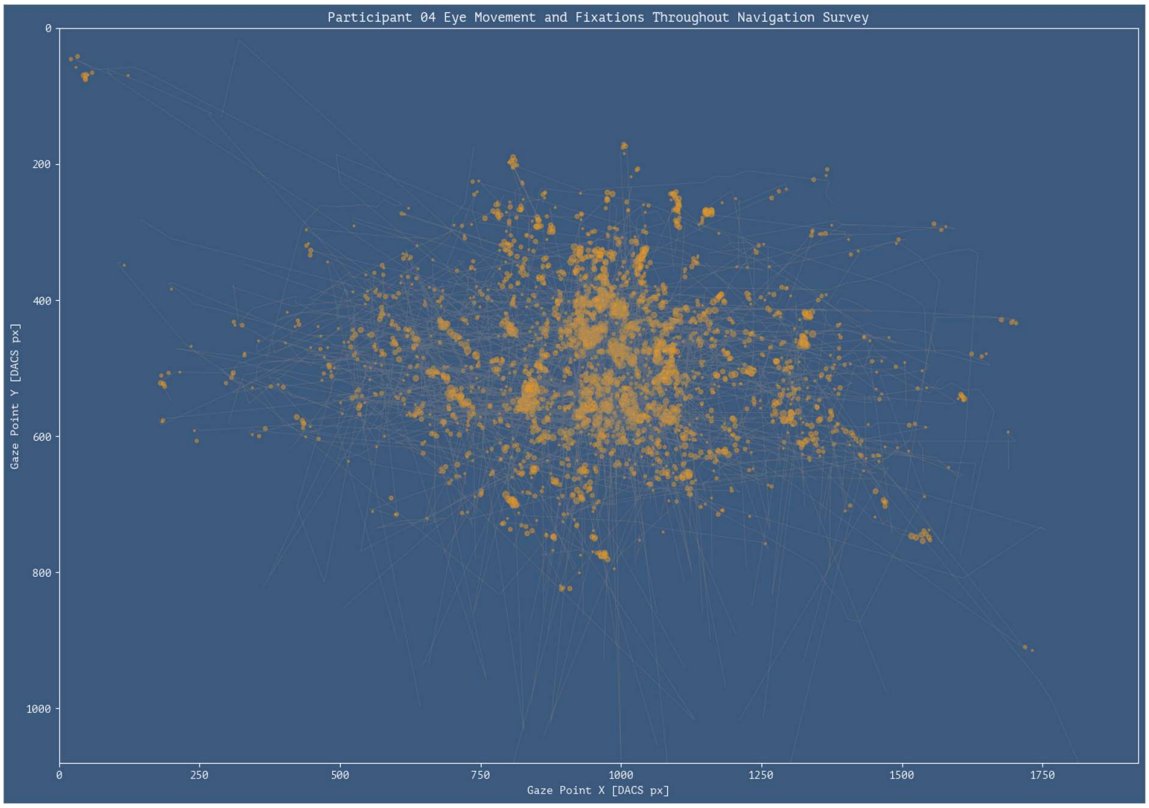

Participant 13's Gaze and Fixation on Objects During Navigation
Object
Fixated Object
Class Name
pyramid
robot
pant
ledge
tower
traffic light
head
balcony
rain
lid
metal sheet
graffiti
glass
fountain
pole
tunnel
hat
balloon
windshield
stack
jacket
bottle
sheet
location
arrow
sunglass
cart
hair
stage
snow
word
water
handle
hole
place
boat
square
computer
toilet
wood
side
saddle
chair
escalator
podium
microwave
house
lot
seat
cat
bridge
fire place
monument
banner
dress
mirror
book
luggage
tie
tire
table
logo
television
net
poster
suitcase
bus station
billboard
roof
sand
hood
helmet
train
clock
painting
paper
statue
bag
retailer
coat
room
blank screen
front
church
box
booth
shirt
image
stair
suit
tent
door
fence
window
sign
computer screen
umbrella
flora
noise
two-wheeler
undetected
street scene
individuals
sky
wall
vehicle
thoroughfare
building
0
1000
2000
3000
4000
5000
6000
Count

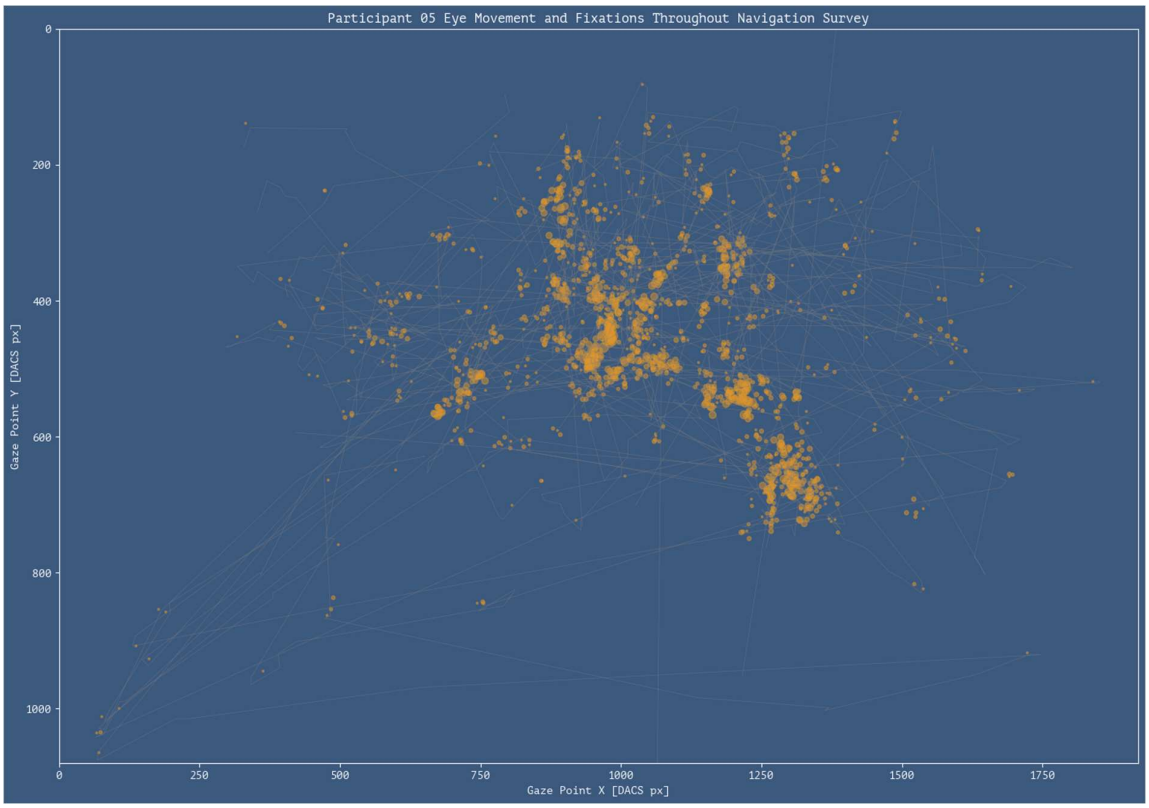
Annex 4: Figures related to Participants Eye movement and Fixation Throughout Navigation Survey

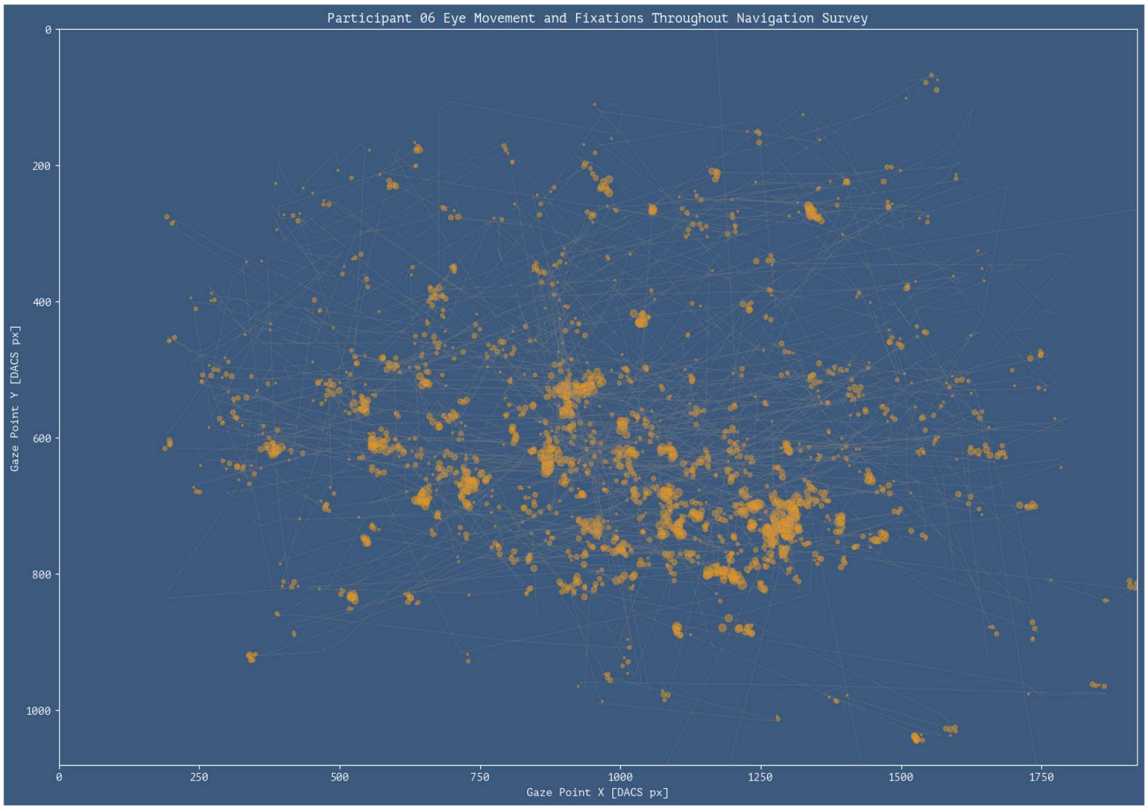


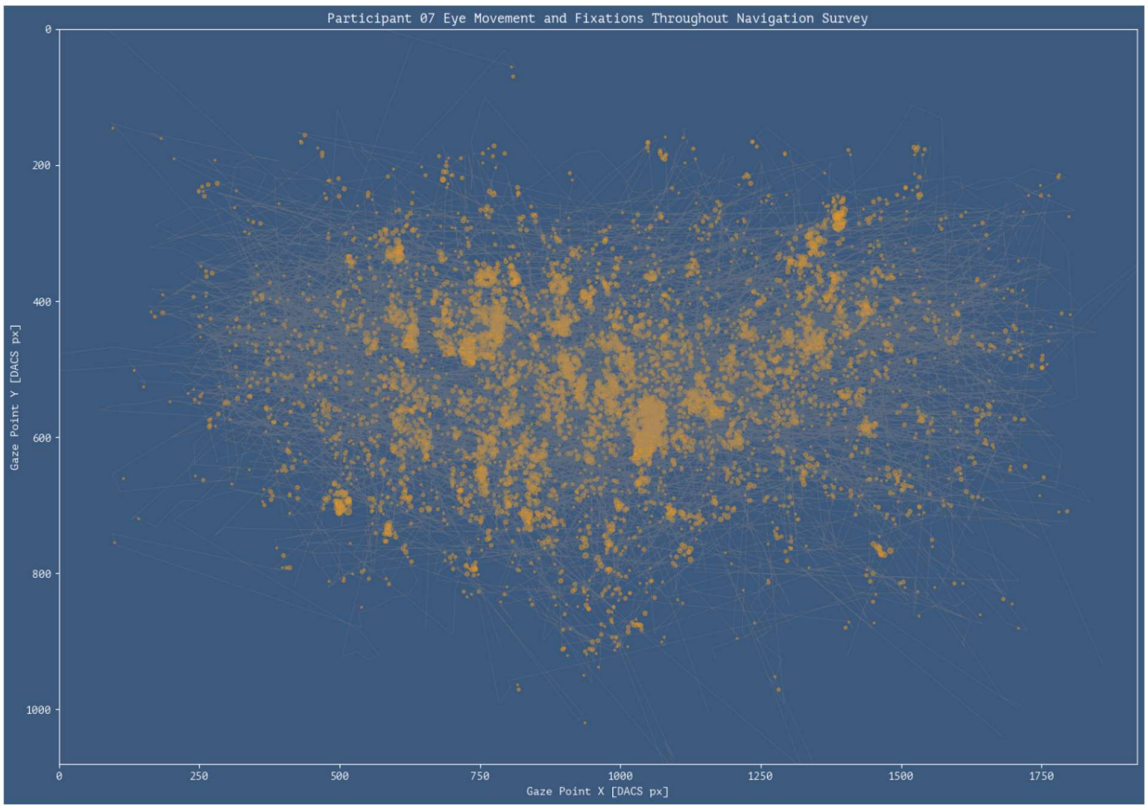

Participant 02 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

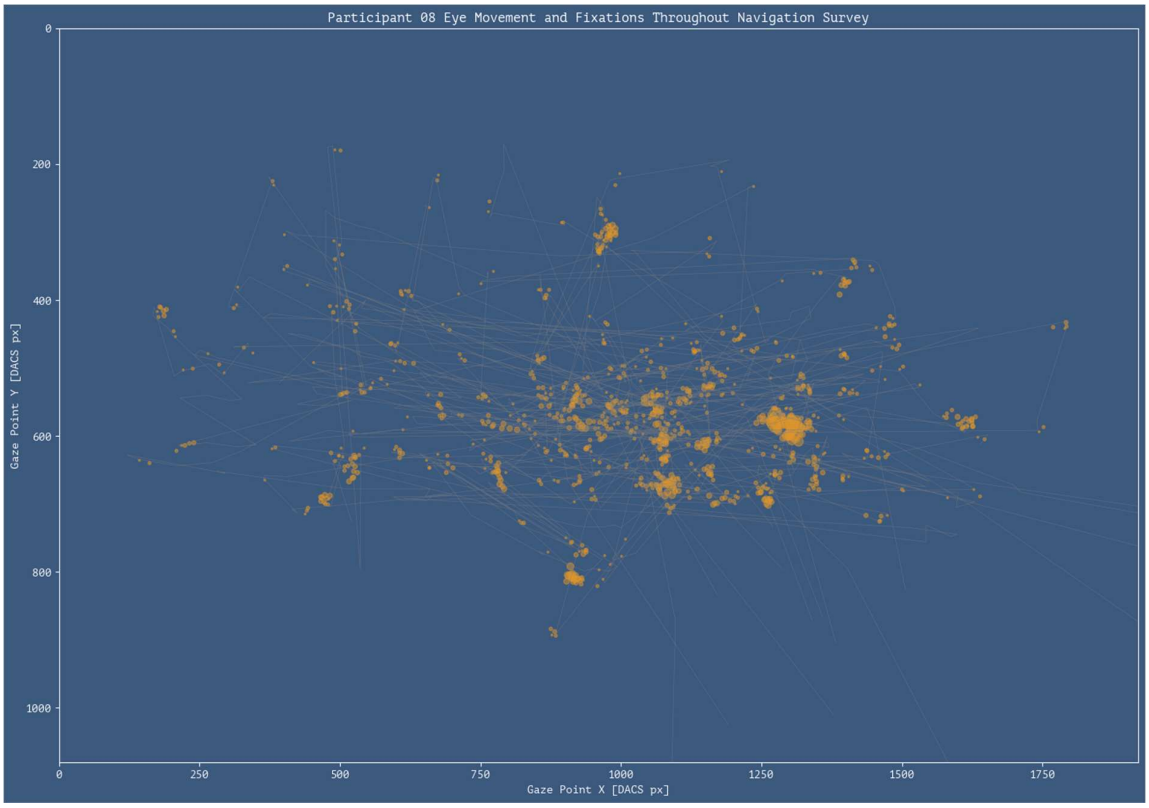

Participant 03 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

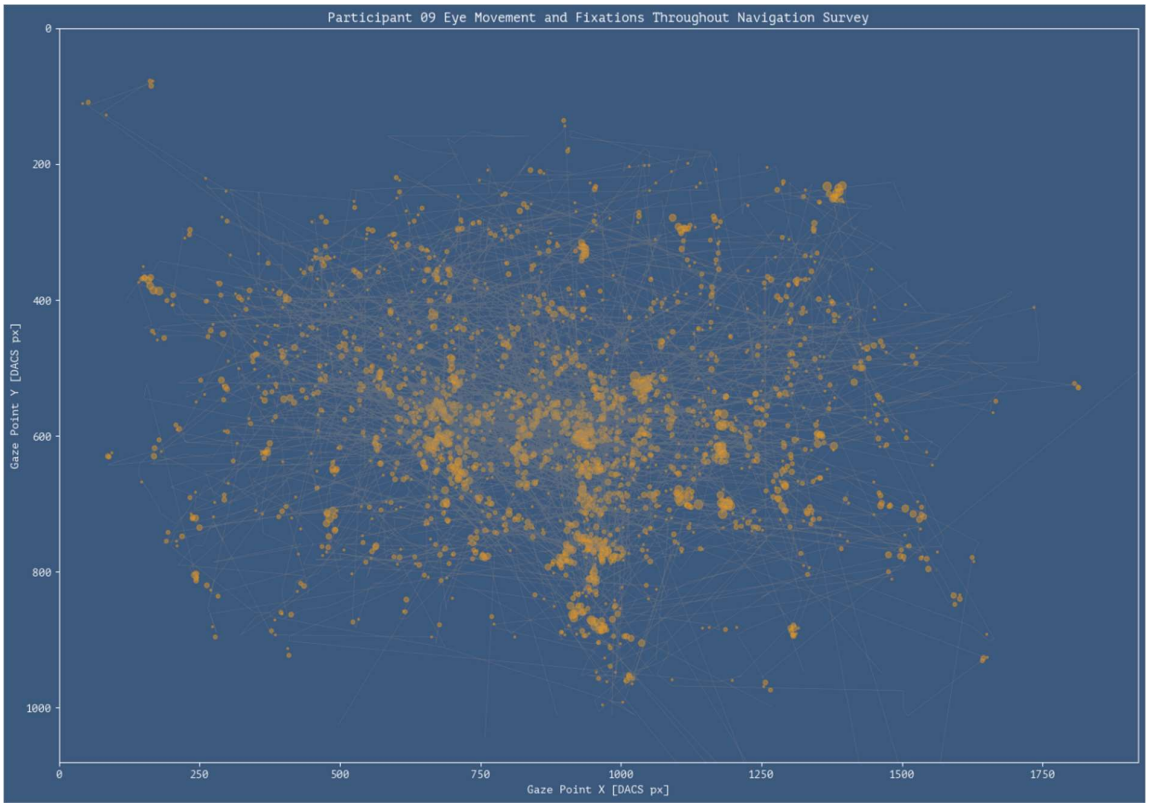

Participant 04 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

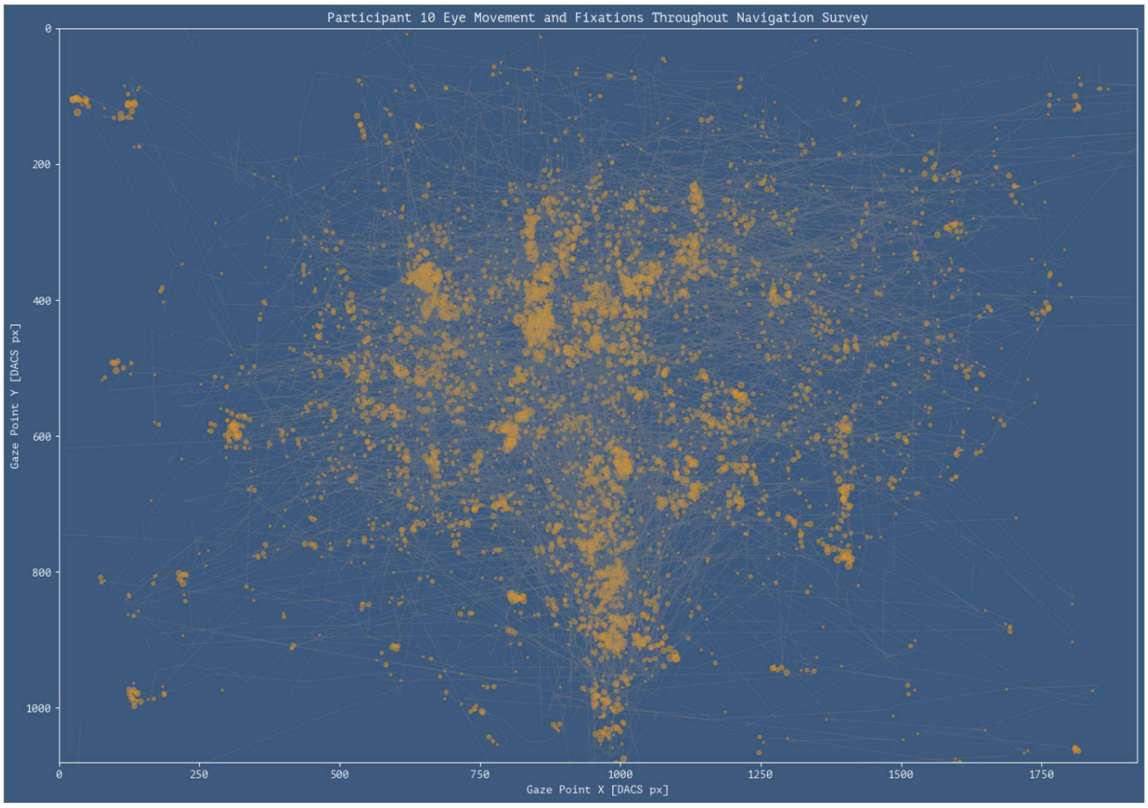

Participant 05 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

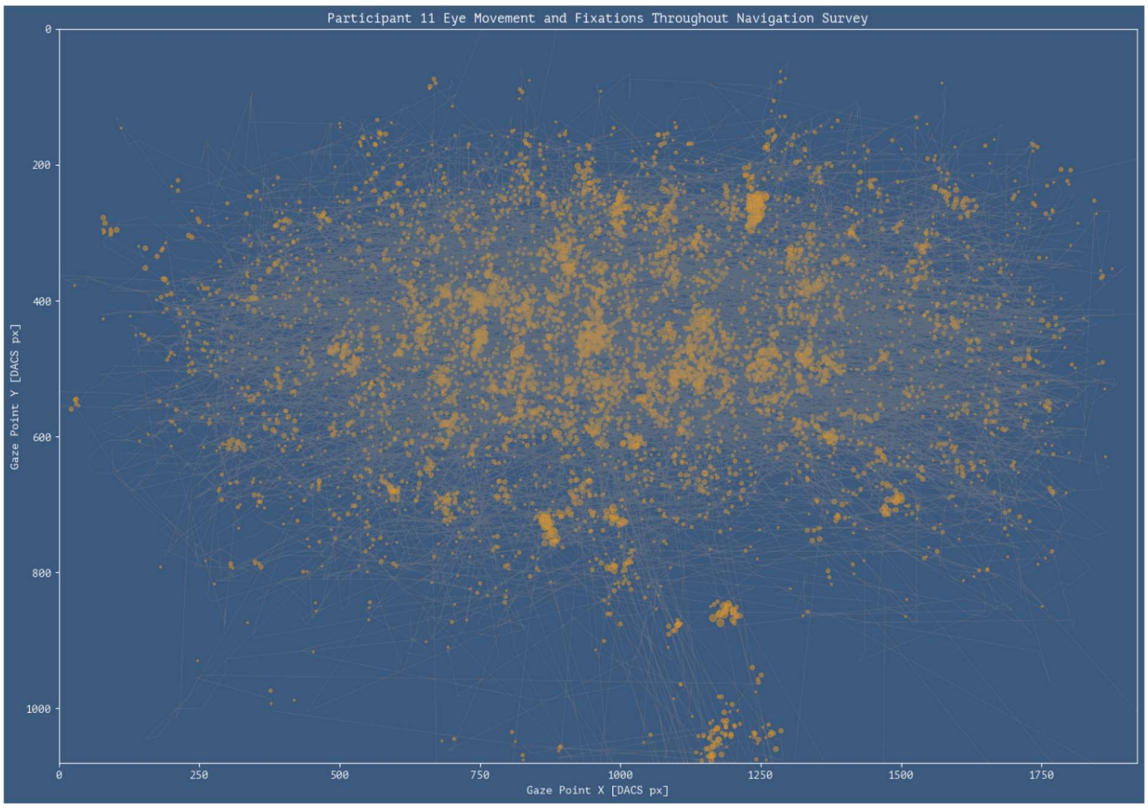

Participant 06 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

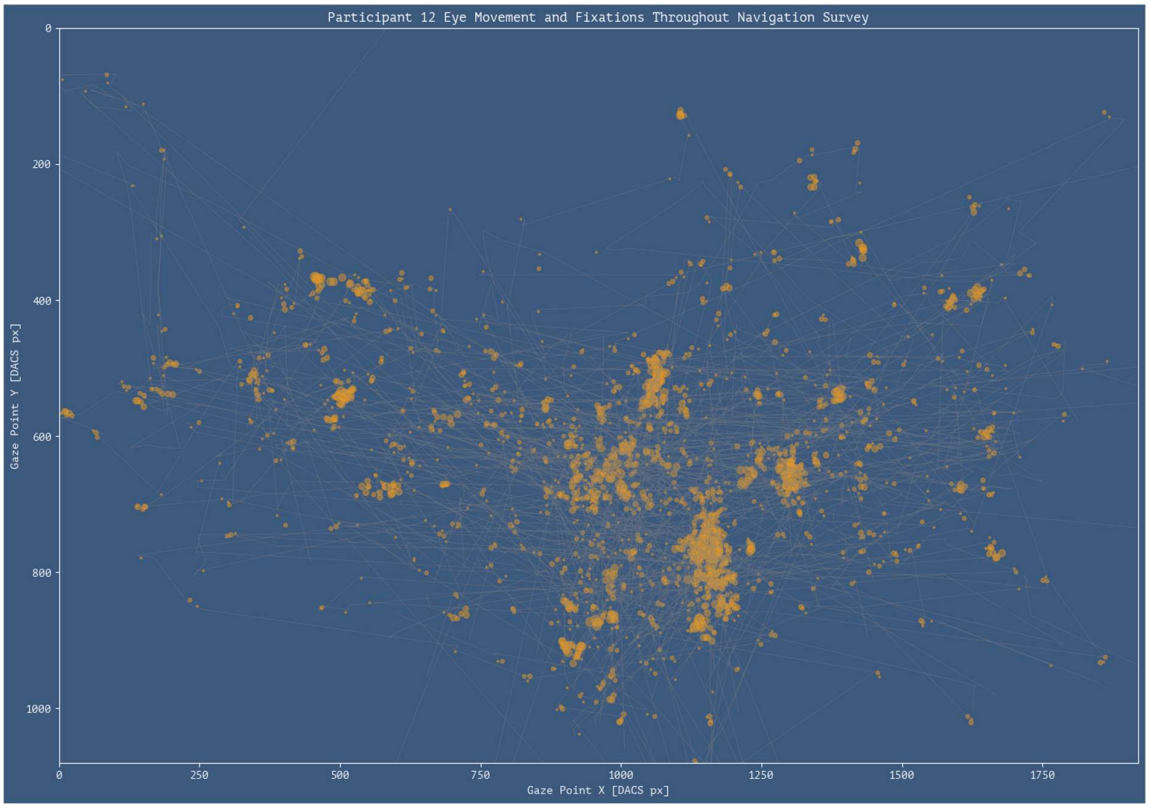

Participant 07 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

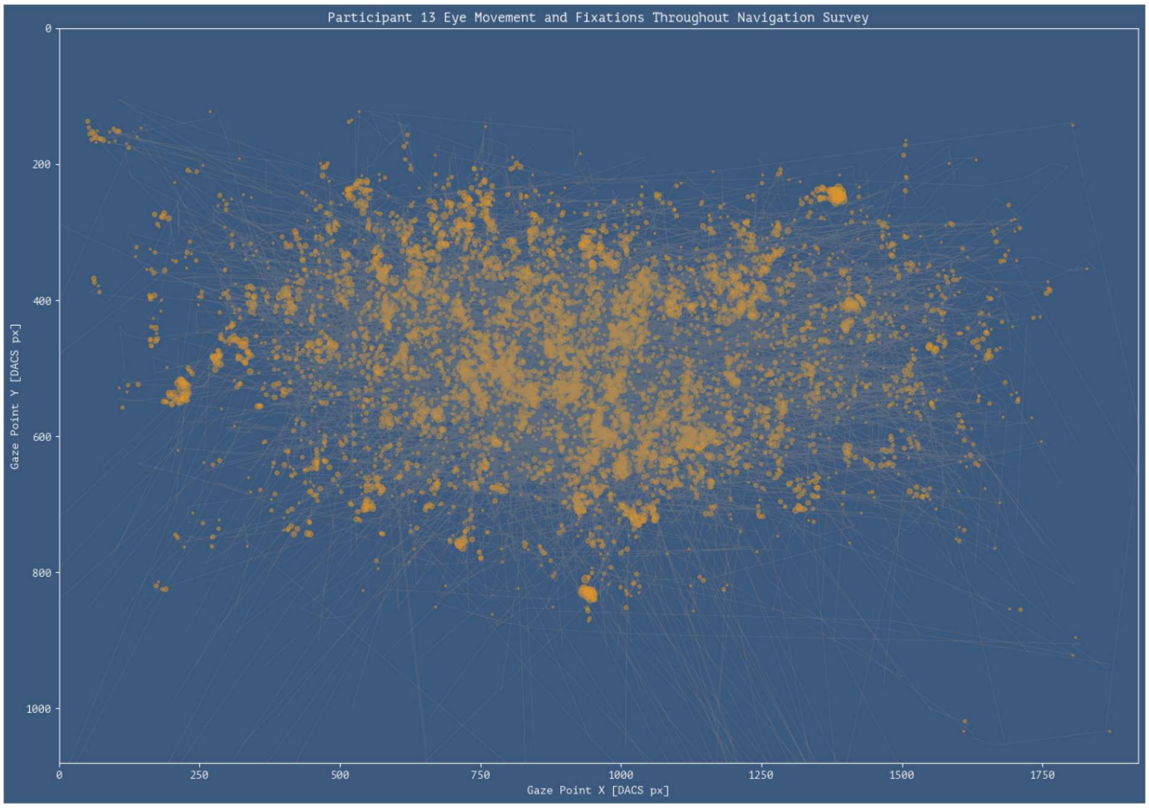

Participant 08 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750


Participant 09 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

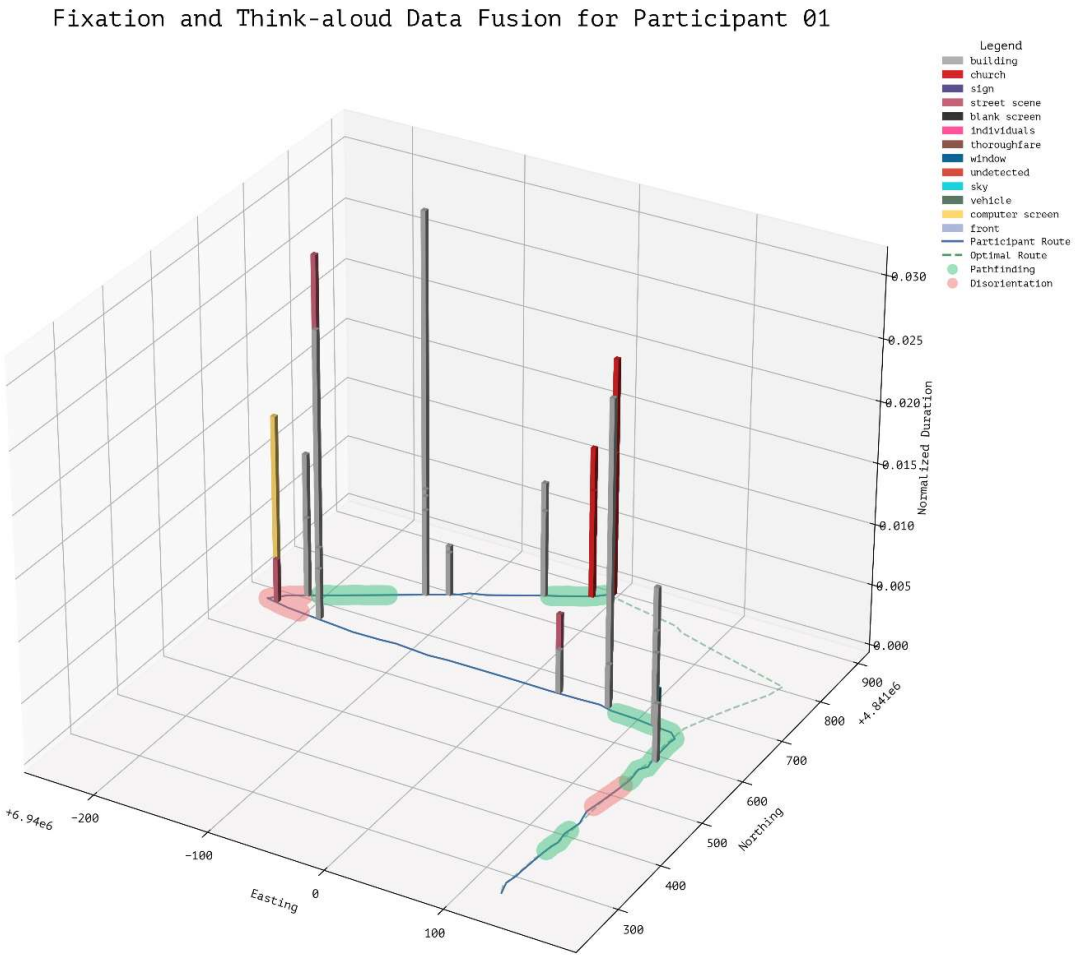

Participant 10 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

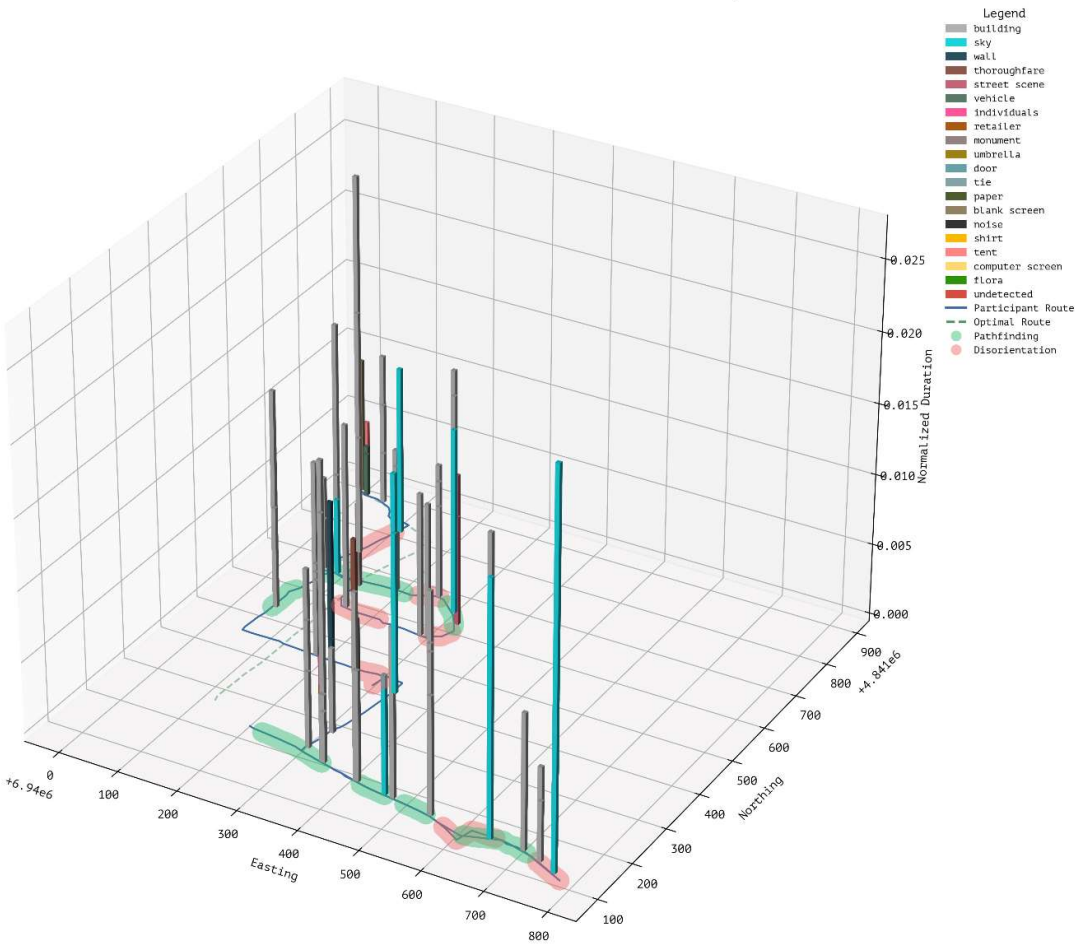

Participant 11 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

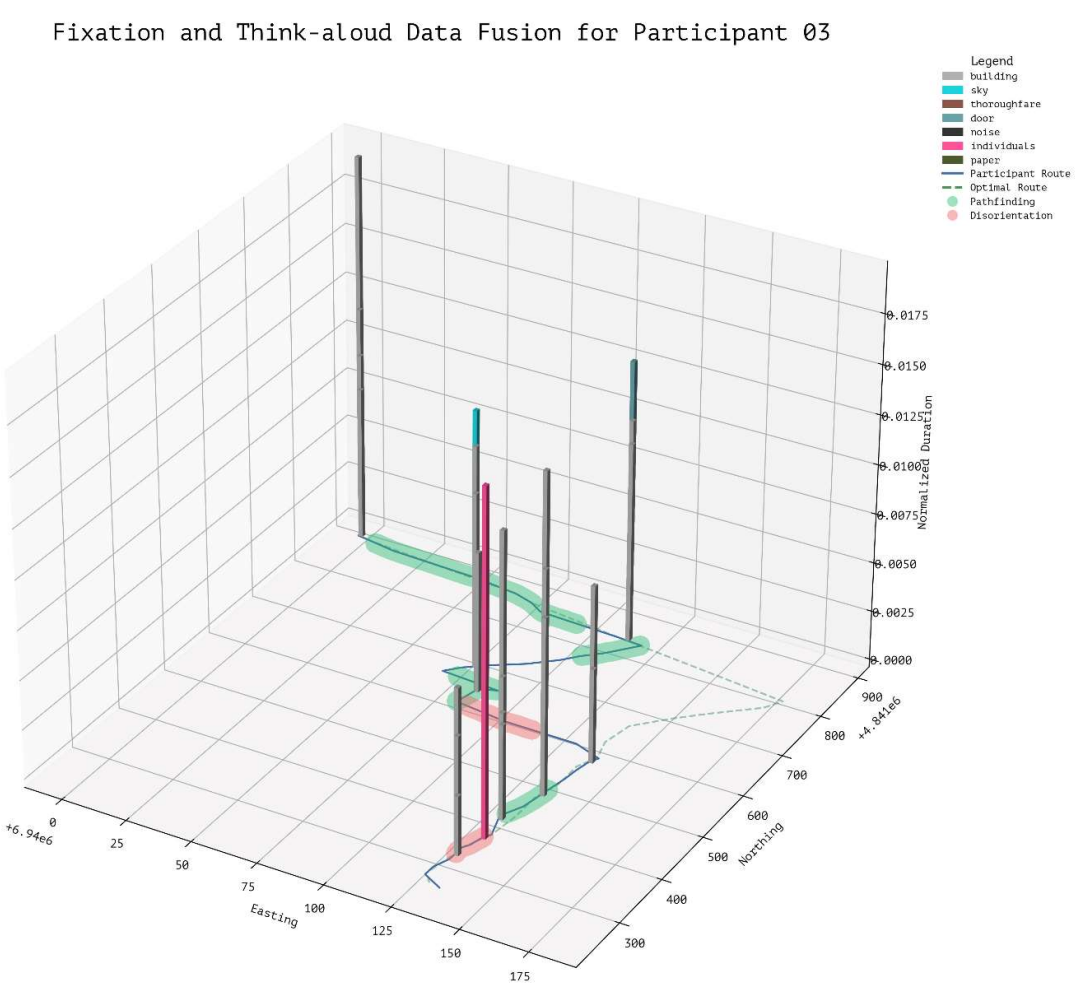

Participant 12 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

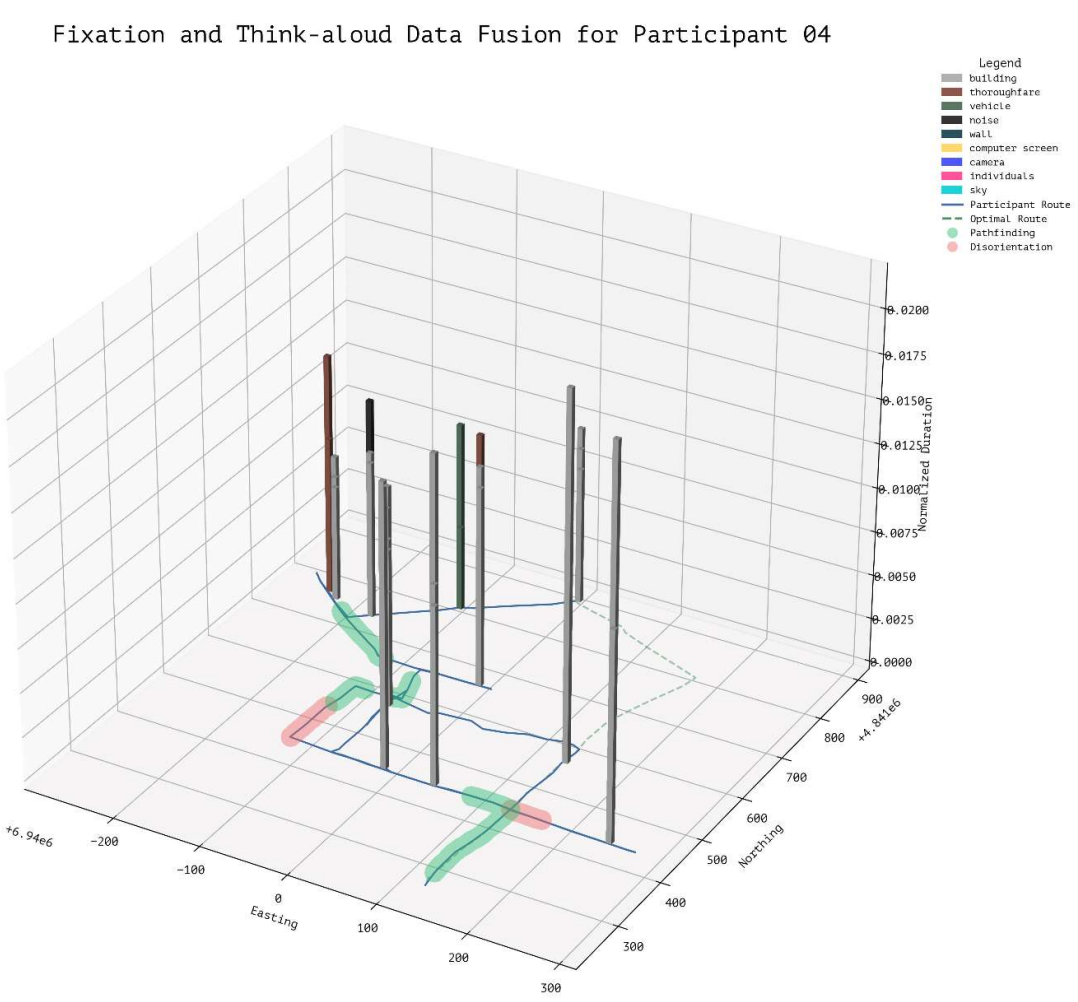

Participant 13 Eye Movement and Fixations Throughout Navigation Survey
Gaze Point Y [DACS px]
Gaze Point X [DACS px]
0
200
400
600
800
1000
0
250
500
750
1000
1250
1500
1750

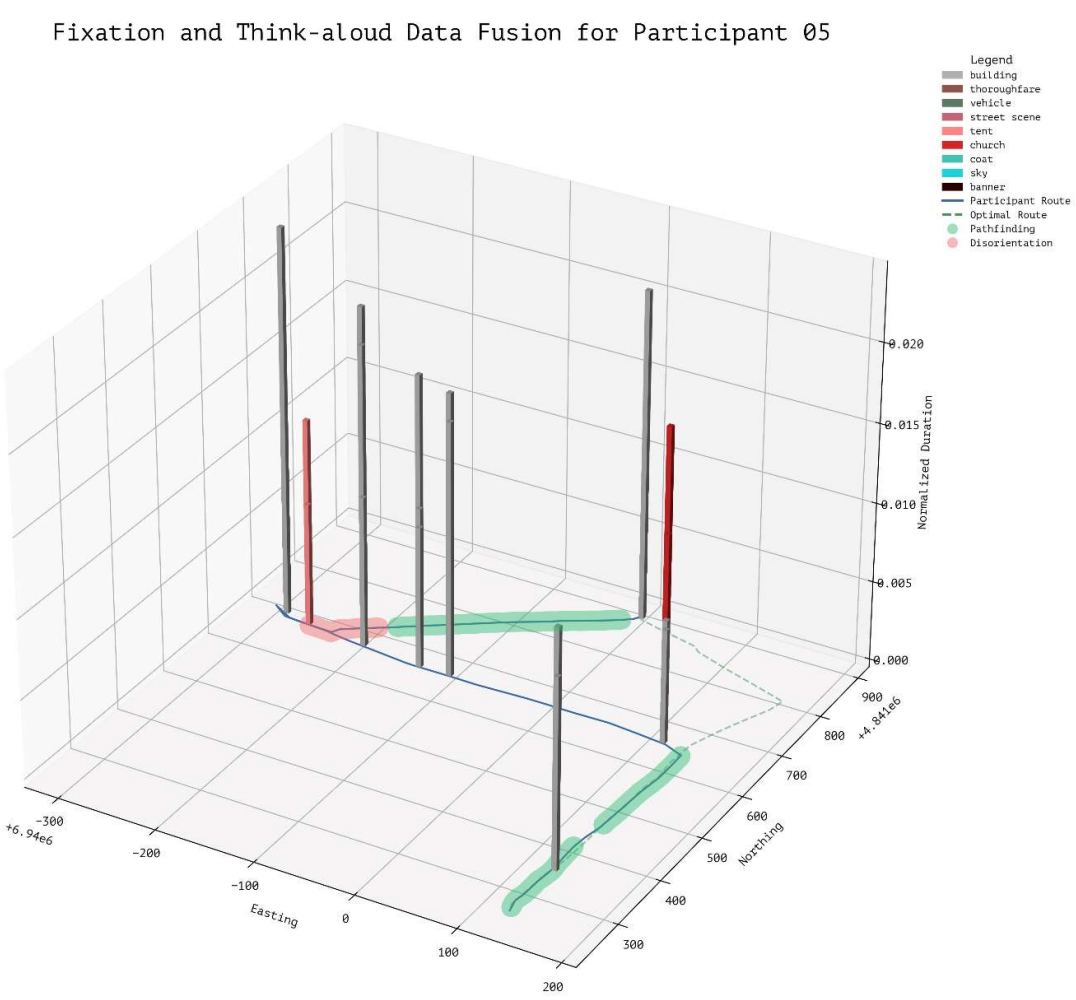
Annex 5: Figures related to Fixation and Think-aloud Data Fusion for Participants

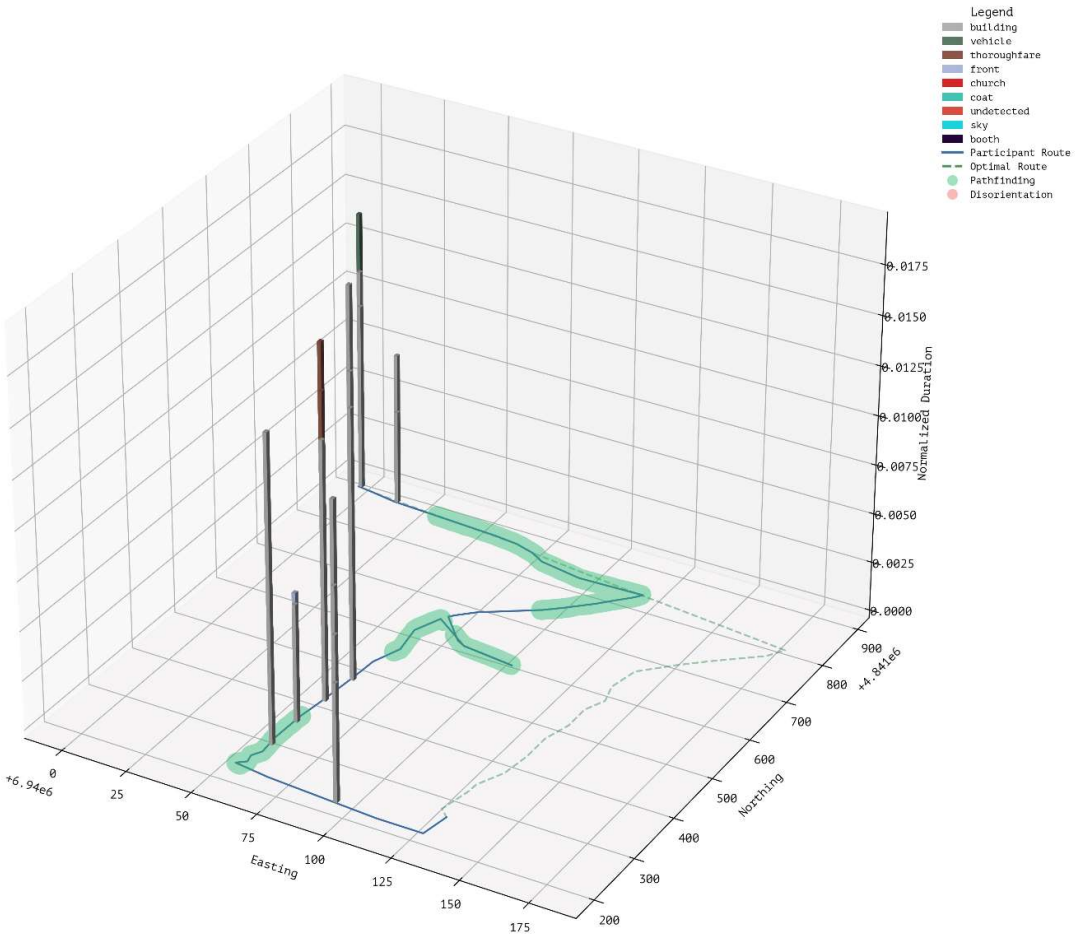


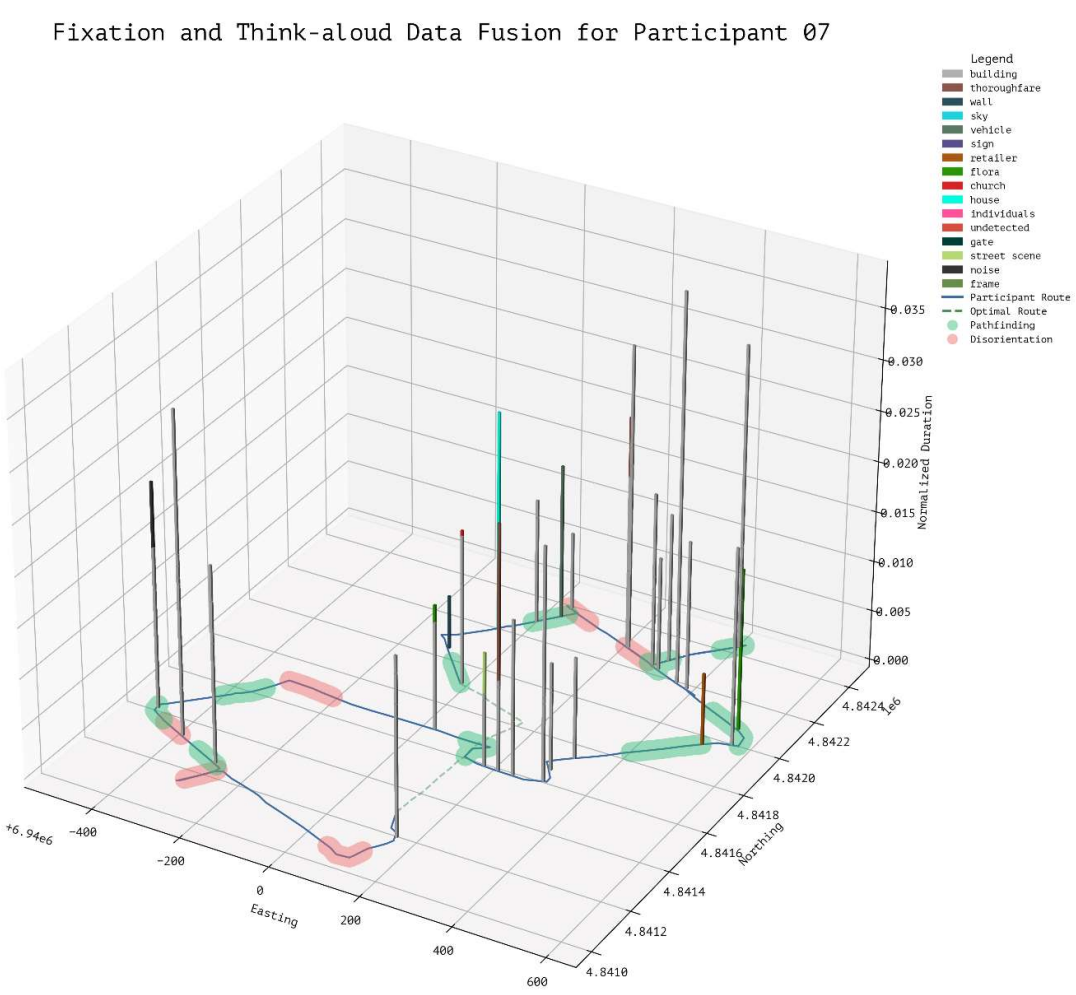

Fixation and Think-aloud Data Fusion for Participant 02
Legend
building
sky
wall
thoroughfare
street scene
vehicle
individuals
retailer
monument
umbrella
door
tie
paper
blank screen
noise
shirt
tent
computer screen
flora
undetected
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.025
0.020
0.015
0.010
0.005
0.000
Easting
+6.94e6
0
100
200
300
400
500
600
700
800
Northing
+4.84e6
100
200
300
400
500
600
700
800
900

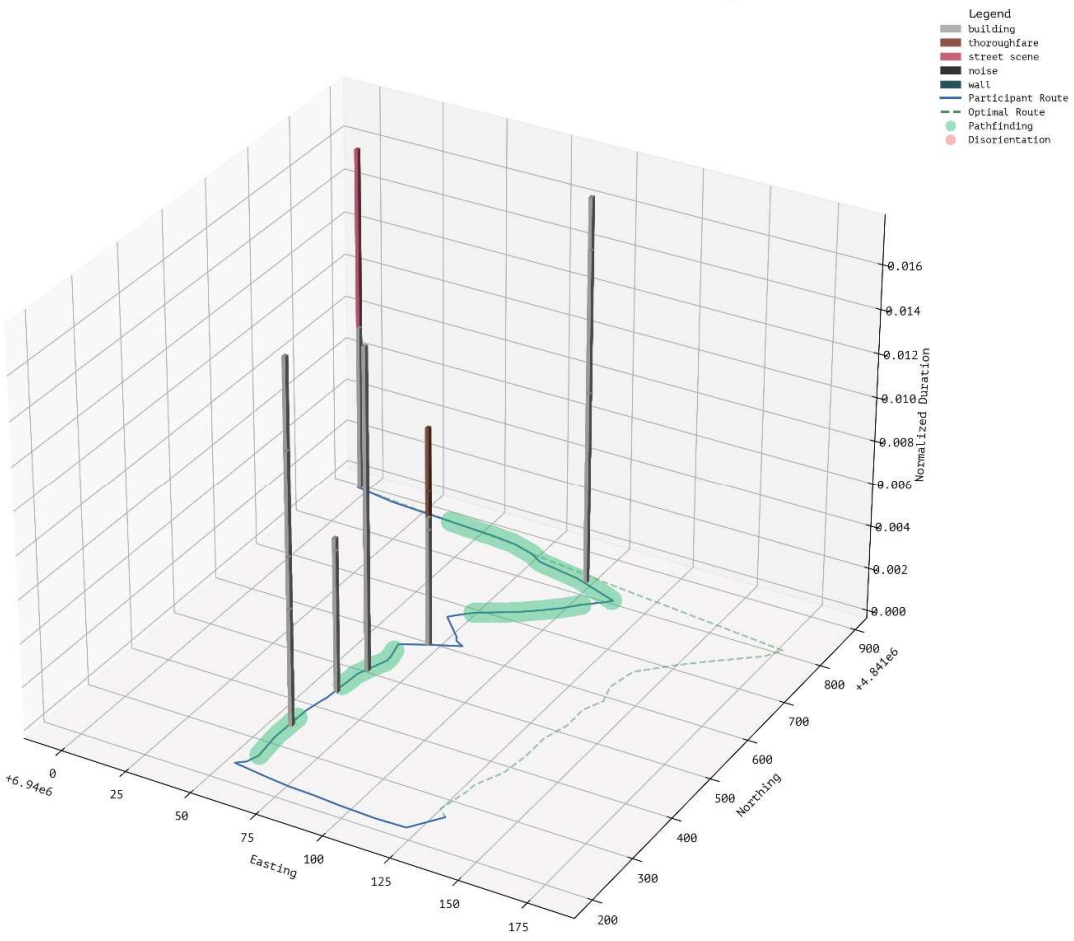

Fixation and Think-aloud Data Fusion for Participant 03
Legend
building
sky
thoroughfare
door
noise
individuals
paper
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
+6.94e6
0
25
50
75
100
125
150
175
Easting
+4.84e6
900
800
700
600
500
400
300
Northing

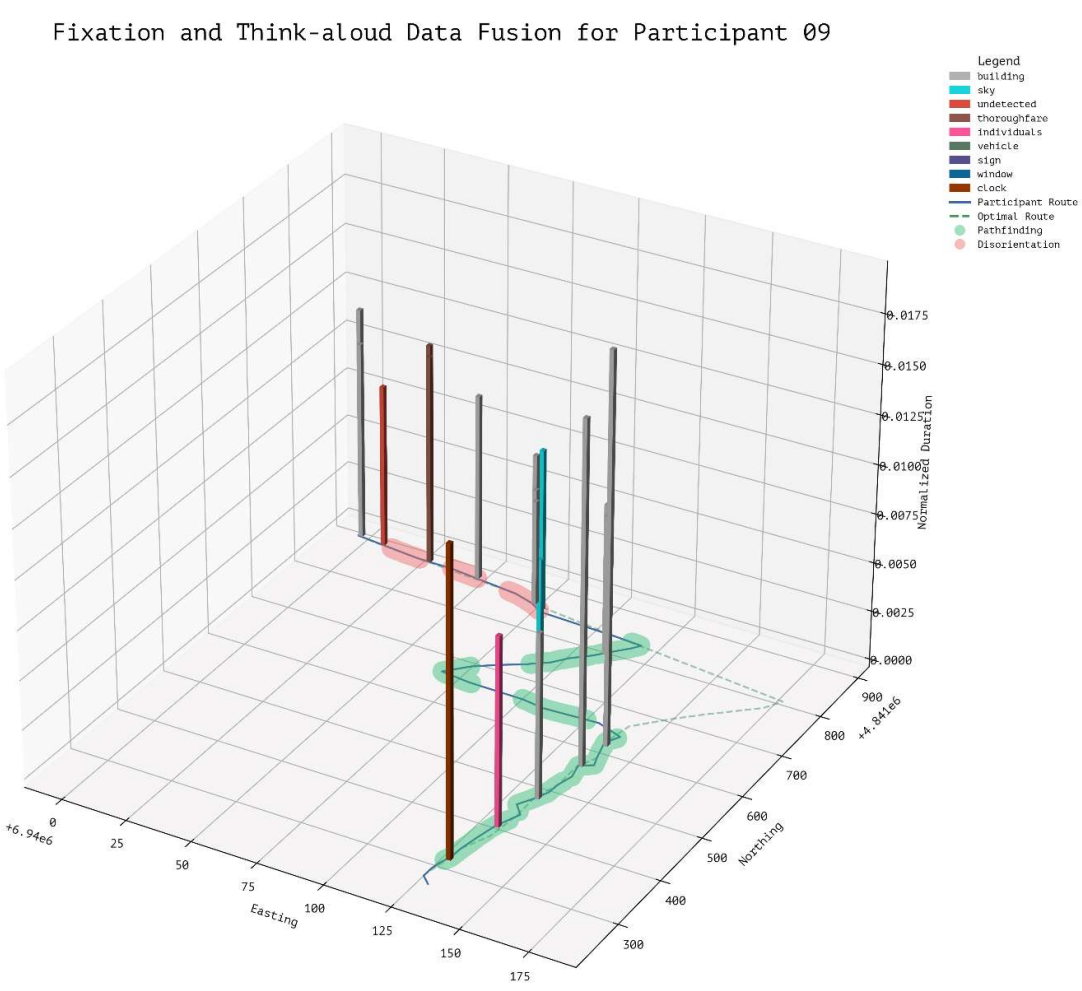

Fixation and Think-aloud Data Fusion for Participant 04
Legend
building
thoroughfare
vehicle
noise
wall
computer screen
camera
individuals
sky
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.0200
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
+6.94e6
−200
−100
0
100
200
300
Easting
300
400
500
600
700
800
900
+4.84e6
Northing

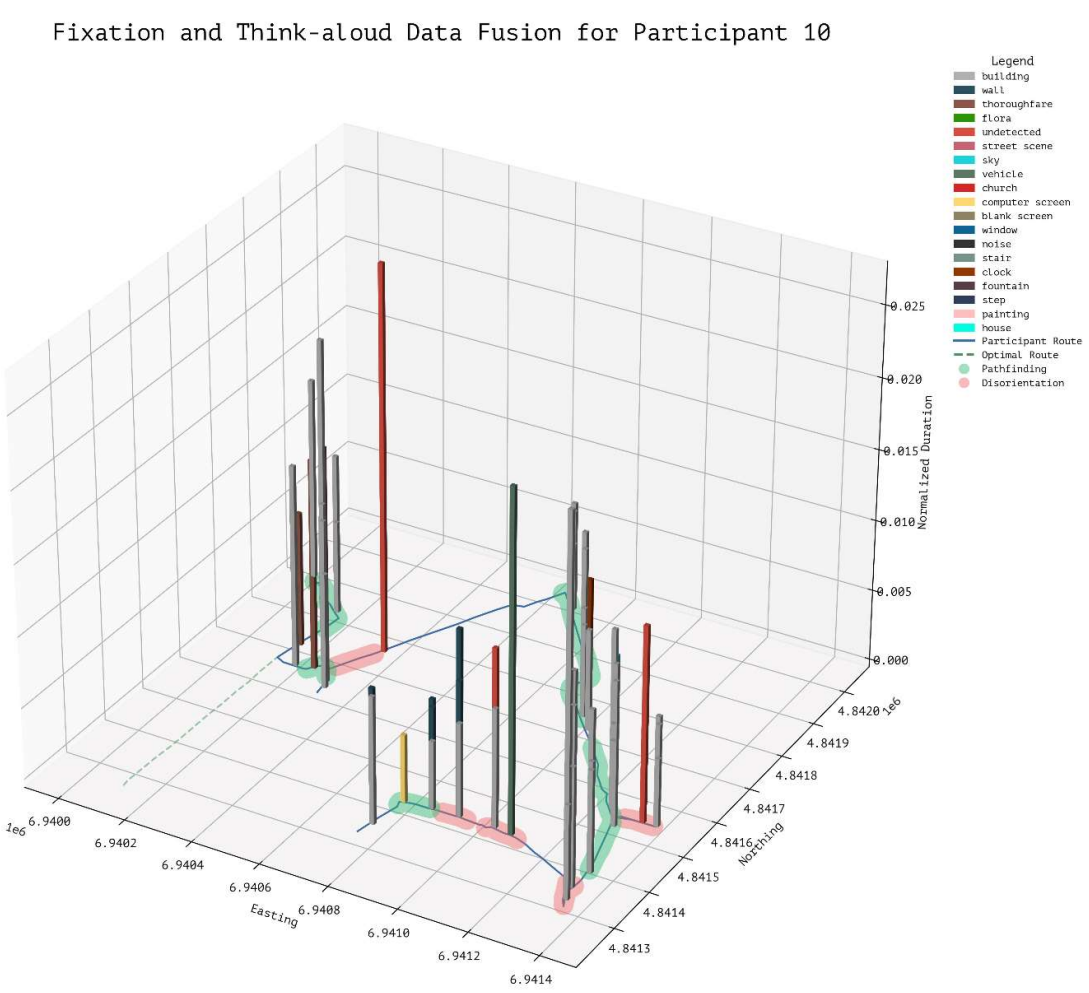

Fixation and Think-aloud Data Fusion for Participant 05
Legend
building
thoroughfare
vehicle
street scene
tent
church
coat
sky
banner
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.020
0.015
0.010
0.005
0.000
+4.8416e6
900
800
700
600
500
400
300
Northing
+6.9466
-300
-200
-100
0
100
200
Easting

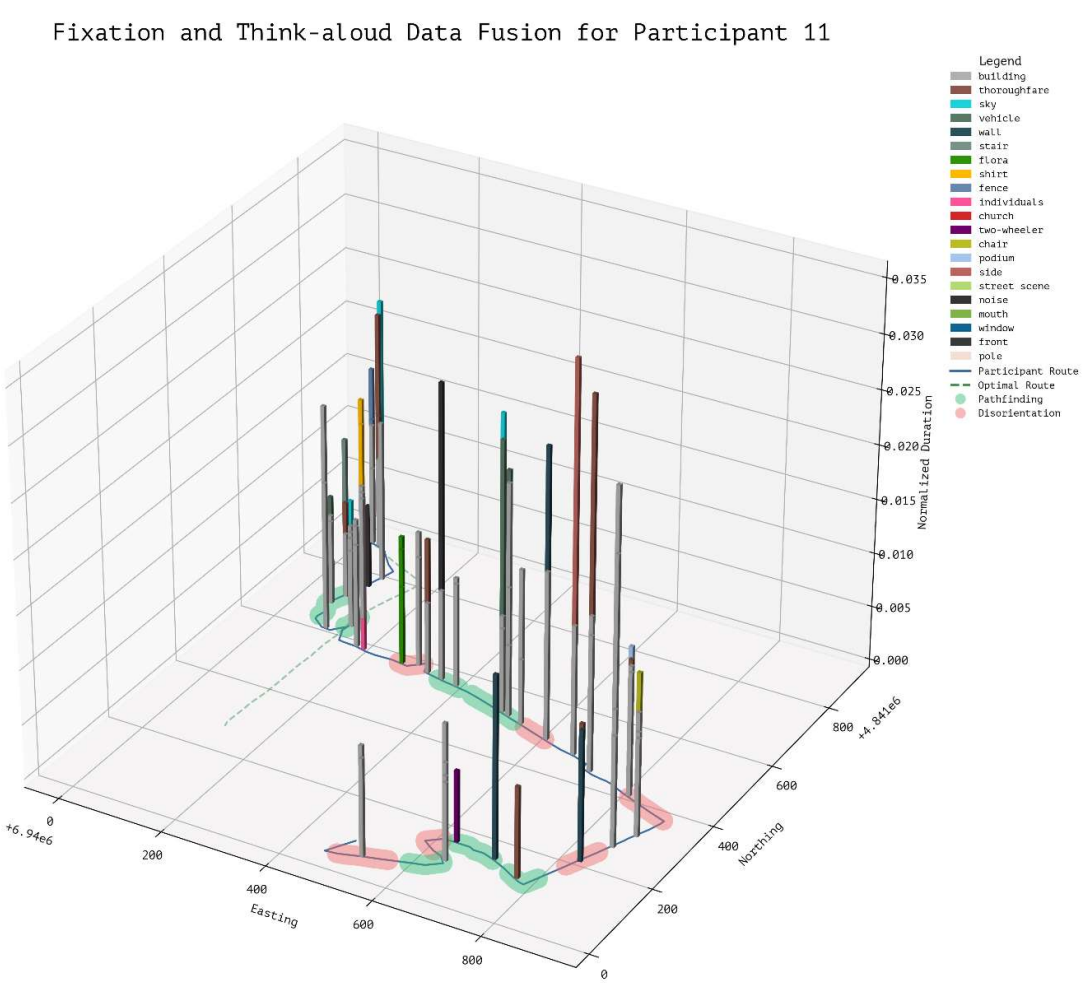

Fixation and Think-aloud Data Fusion for Participant 06
Legend
building
vehicle
thoroughfare
front
church
coat
undetected
sky
booth
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
+6.94e6
0
25
50
75
100
125
150
175
Easting
+4.84e6
900
800
700
600
500
400
300
200
Northing

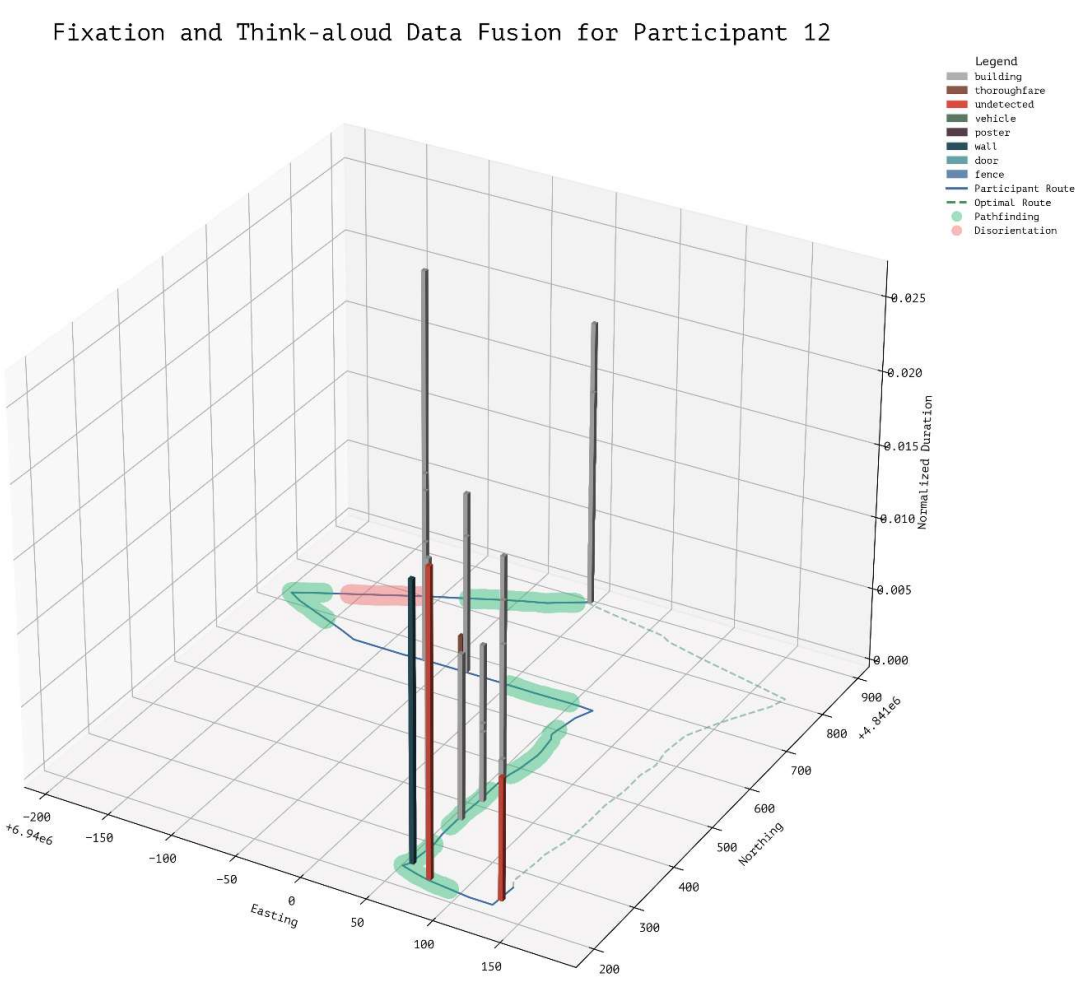

Fixation and Think-aloud Data Fusion for Participant 07
Legend
building
thoroughfare
wall
sky
vehicle
sign
retailer
flora
church
house
individuals
undetected
gate
street scene
noise
frame
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
+6.94e6
-400
-200
0
200
400
600
Easting
4.8424
+6
4.8422
4.8420
4.8418
4.8416
4.8414
4.8412
4.8410
Northing

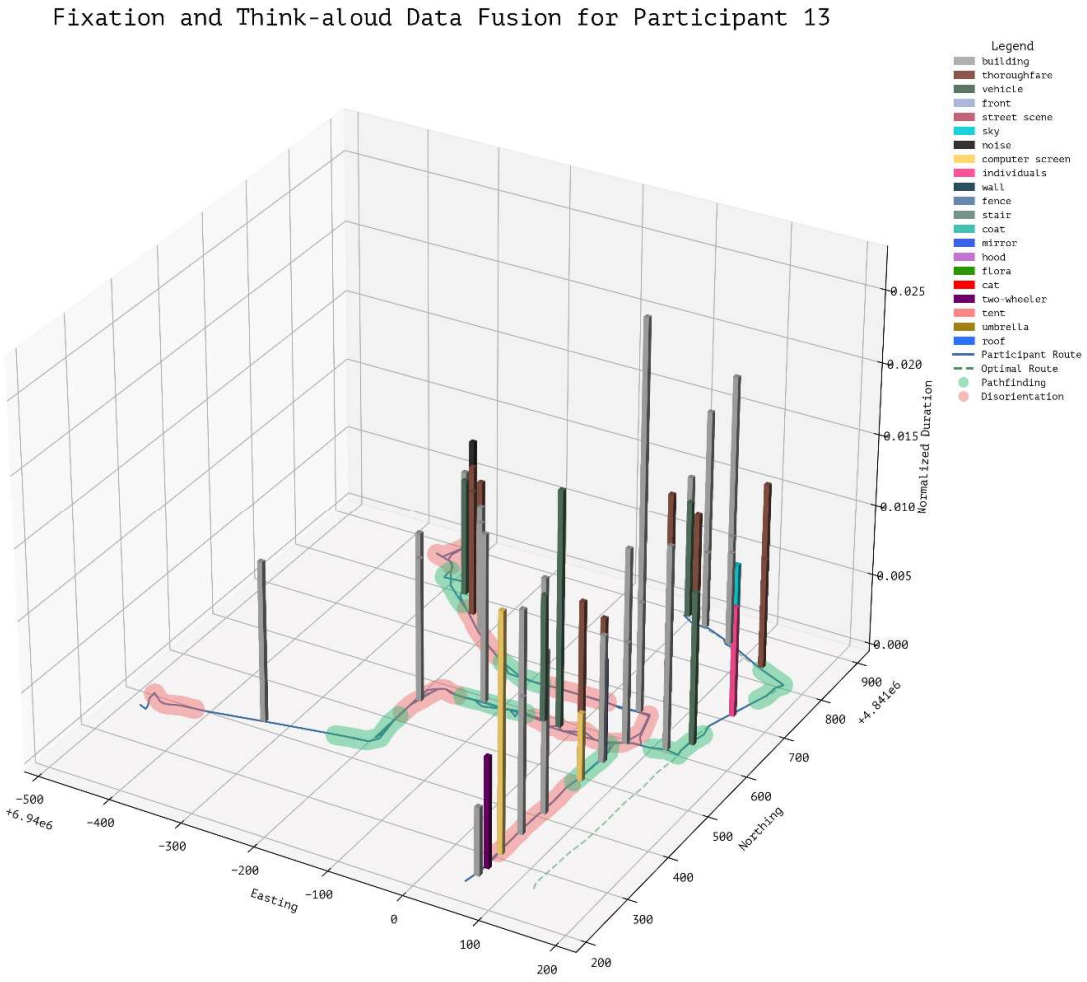

Fixation and Think-aloud Data Fusion for Participant 08
Legend
building
thoroughfare
street scene
noise
wall
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
+4.84e6
900
800
700
600
500
400
300
200
Northing
+6.94e6
0
25
50
75
100
125
150
175
Easting


Fixation and Think-aloud Data Fusion for Participant 09
Legend
building
sky
undetected
thoroughfare
individuals
vehicle
sign
window
clock
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
+4.841e6
900
800
700
600
500
400
300
Northing
+6.94e6
0
25
50
75
100
125
150
175
Easting


Fixation and Think-aloud Data Fusion for Participant 10
Legend
building
wall
thoroughfare
flora
undetected
street scene
sky
vehicle
church
computer screen
blank screen
window
noise
stair
clock
fountain
step
painting
house
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.025
0.020
0.015
0.010
0.005
0.000
1e6
6.9400
6.9402
6.9404
6.9406
6.9408
6.9410
6.9412
6.9414
Easting
1e6
4.8420
4.8419
4.8418
4.8417
4.8416
4.8415
4.8414
4.8413
Northing


Fixation and Think-aloud Data Fusion for Participant 11
Legend
building
thoroughfare
sky
vehicle
wall
stair
flora
shirt
fence
individuals
church
two-wheeler
chair
podium
side
street scene
noise
mouth
window
front
pole
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
+6.94e6
0
200
400
600
800
Easting
+4.841e6
0
200
400
600
800
Northing


Fixation and Think-aloud Data Fusion for Participant 12
Legend
building
thoroughfare
undetected
vehicle
poster
wall
door
fence
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.025
0.020
0.015
0.010
0.005
0.000
Easting
+6.94e6
−200
−150
−100
−50
0
50
100
150
Northing
+4.84e6
200
300
400
500
600
700
800
900


Fixation and Think-aloud Data Fusion for Participant 13
Legend
building
thoroughfare
vehicle
front
street scene
sky
noise
computer screen
individuals
wall
fence
stair
coat
mirror
hood
flora
cat
two-wheeler
tent
umbrella
roof
Participant Route
Optimal Route
Pathfinding
Disorientation
Normalized Duration
0.025
0.020
0.015
0.010
0.005
0.000
Easting
+6.94e6
-500
-400
-300
-200
-100
0
100
200
Northing
+4.84e6
900
800
700
600
500
400
300
200